

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-191394

(43) 公開日 平成9年(1997)7月22日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 N 1/387			H 0 4 N 1/387	
G 0 6 T 1/00		7259-5 J	G 0 9 C 5/00	
G 0 9 C 5/00			G 1 0 L 3/00	M
G 1 0 L 3/00			G 0 6 F 15/66	B

審査請求 有 請求項の数30 OL 外国語出願 (全 53 頁)

(21) 出願番号 特願平8-251801  
 (22) 出願日 平成8年(1996)9月24日  
 (31) 優先権主張番号 08/534894  
 (32) 優先日 1995年9月28日  
 (33) 優先権主張国 米国 (US)

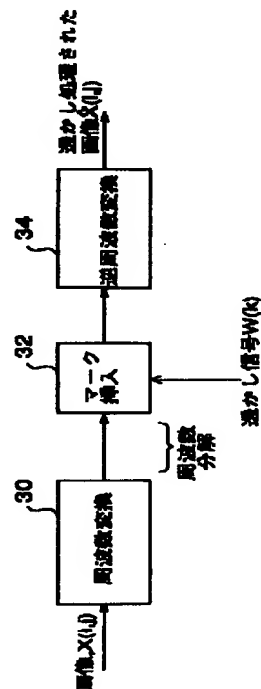
(71) 出願人 000004237  
 日本電気株式会社  
 東京都港区芝五丁目7番1号  
 (72) 発明者 インゲマー コックス  
 アメリカ合衆国, ニュージャージー  
 08648, ローレンスヴィル, レ パーク  
 ドライヴ 21  
 (72) 発明者 ジョセフ キリアン  
 アメリカ合衆国, ニュージャージー  
 08550, プリンストン ジャンクション,  
 リード ドライヴ ノース 18  
 (74) 代理人 弁理士 後藤 洋介 (外2名)

最終頁に続く

(54) 【発明の名称】 電子的すかし挿入方法

## (57) 【要約】

オーディオ、画像、映像、あるいはマルチメディアデータの電子透かし処理が、視覚的には知覚できないような方法で、分解されたデータの知覚的に重要な成分に、透かしの挿入することによって達成される。好ましい方法においては、データの周波数スペクトラム画像、好ましくはデータのフーリエ変換によって達成される。透かしは、周波数スペクトラム画像の知覚的に重要な成分に挿入される。結果的に透かし処理されたスペクトラム画像は逆変換を施されることによって、透かし処理されたデータが生成される。透かしは、透かし処理されたデータとオリジナルのデータとを最初に比較して、透かしの抽出することによって、透かし処理されたデータから抽出される。それから、オリジナルの透かし、オリジナルデータ、および抽出された透かしが比較され、透かしの真正さが分析される。



## 【特許請求の範囲】

【請求項1】 透かしが施されるデータを分解するステップ、分解されたデータのうちの知覚的に重要な成分中に透かしを挿入するステップ、および、透かしが施された分解されたデータに逆変換を施し、透かし処理されたデータを生成するステップを有することを特徴とするデータ中に透かしを挿入する方法。

【請求項2】 請求項1において、前記データは画像データを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項3】 請求項1において、前記データは映像データを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項4】 請求項1において、前記データはオーディオデータを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項5】 請求項1において、前記データはマルチメディアデータを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項6】 請求項1において、前記透かしを挿入するステップは、データの知覚的品質に影響を受けるような知覚的に重要な成分中へ付加的信号を付加することによって透かし値を挿入するステップを含んでいることを特徴とするデータ中に透かしを挿入する方法。

【請求項7】 請求項1において、前記分解されたデータを得るステップは、データをスペクトラム分解することによって、分解されたデータを得ることを特徴とするデータ中に透かしを挿入する方法。

【請求項8】 請求項7において、前記データは画像データを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項9】 請求項7において、前記データは映像データを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項10】 請求項7において、前記データはオーディオデータを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項11】 請求項7において、前記データはマルチメディアデータを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項12】 請求項7において、スペクトラム分解されたデータを得る前記ステップは、フーリエ変換、離散余弦変換、アダマル変換、およびウェーブレット、多重分解能、サブバンド方法から成るグループから選ばれていることを特徴とするデータ中に透かしを挿入する方法。

【請求項13】 請求項12において、前記透かしを挿入するステップは、データの知覚的品質に影響を与えるような知覚的に重要な成分中へ付加的信号を付加することによって透かし値を挿入するステップを含んでいるこ

とを特徴とするデータ中に透かしを挿入する方法。

【請求項14】 請求項7において、さらに、データと透かしが施されたデータとを比較して抽出されたデータ値を得るステップ、すかし値を有する抽出されたデータ値とデータとを比較して、両者の差を得るためのステップ、および、前記差を分析して、透かしが施されたデータ中の透かしを決定するステップを有することを特徴とするデータ中に透かしを挿入する方法。

【請求項15】 請求項14において、透かし値は、関連したスケーリングパラメータを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項16】 請求項15において、スケーリングパラメータは、データの知覚的品質に影響を与えるような付加的な透かし値を付加するようなスケーリングパラメータから選ばれることを特徴とするデータ中に透かしを挿入する方法。

【請求項17】 請求項14において、透かし値は、正規分布に従って選ばれることを特徴とするデータ中に透かしを挿入する方法。

【請求項18】 スペクトラム分解したデータの知覚的に重要な成分の値を抽出するステップ、透かし値と抽出された値とを結合して調整された値を生成するステップ、および抽出された値の代わりにデータ中へ調整された値を挿入し、透かし処理されたデータを得るステップを有することを特徴とするデータ中に透かしを挿入する方法。

【請求項19】 請求項18において、透かし値は、関連したスケーリングパラメータを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項20】 請求項19において、スケーリングパラメータは、データの知覚的品質に影響を与えるような付加的な透かし値を付加するスケーリングパラメータから選ばれることを特徴とするデータ中に透かしを挿入する方法。

【請求項21】 請求項18において、透かし値は、ランダム分布に応じて選ばれることを特徴とするデータ中に透かしを挿入する方法。

【請求項22】 請求項18において、さらに、データと透かし処理されたデータとを比較し、抽出されたデータ値を得るステップ、抽出されたデータを透かし処理されたデータおよびデータと比較し、差値を得るステップ、および、差値を分析して、透かしが施されたデータ中の透かしを決定するステップを有することを特徴とするデータ中に透かしを挿入する方法。

【請求項23】 請求項22において、透かし値は、関連するスケーリングパラメータを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項24】 請求項23において、スケーリングパラメータは、データの知覚的品質に影響を与えるような付加的な透かし値を付加するスケーリングパラメータか

ら選ばれることを特徴とするデータ中に透かしを挿入する方法。

【請求項25】 請求項22において、透かし値は、ランダム分布に応じて選ばれることを特徴とするデータ中に透かしを挿入する方法。

【請求項26】 請求項22において、さらに、前記データを比較するステップの前に歪み処理またはタンバリング処理され、且つ、透かし処理されたデータを前処理するステップを有することを特徴とするデータ中に透かしを挿入する方法。

【請求項27】 請求項26において、前記歪みまたはタンバリング処理され、且つ透かし処理されたデータは、クリップされたデータであると共に、前記前処理はデータ中の消失部分を、透かしの施されていないオリジナルデータの対応部分と置き替えるステップを含むことを特徴とするデータ中に透かしを挿入する方法。

【請求項28】 画像データおよび透かし画像データを与える手段と、画像データを通過させ、変換された画像データに変換する第1の変換レンズと、透かし画像データを通過させ、変換された透かし画像データに変換する第2の変換レンズと、変換され、且つ透かし処理されたデータを形成するために、変換された画像データおよび変換された透かし画像データを結合する光学的結合器と、変換され、且つ透かし処理されたデータを逆変換して、透かし処理データを形成する逆変換レンズを有することを特徴とするデータ中に透かしを挿入するシステム。

【請求項29】 請求項28において、前記第1および第2の変換レンズはフーリエ変換レンズであり、かつ前記逆変換レンズは逆フーリエ変換レンズであることを特徴とするデータ中に透かしを挿入するシステム。

【請求項30】 透かしを施すべき分解されたデータを得るステップ、データに歪みおよび/あるいはタンバリング処理を施すことによって、透かし処理されるべきデータを修正するステップ、分解された修正データを得るステップ、透かしを施すべき分解されたデータの成分と、分解された修正データの成分とを比較するステップ、および前記比較結果に基づいて透かしを施すべきデータ中に透かしを挿入するステップを有することを特徴とするデータ中に透かしを挿入する方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、著作権の所有権を証明し、著作権侵害を識別し、隠されたメッセージを送信するために、オーディオ、画像、ビデオ、マルチメディアデータに使用される電子透かし（デジタル透かし、即ち、ウォーターマーク）に関する。特に、本発明に係る透かしは、データを分解した中で知覚的に最も重要な成分に、実際視覚的には感知できないような方法で挿入される。より具体的に言えば、透かしを表す狭帯域の信

号は、データが存在する広帯域チャンネル内に配置される。

【0002】

【従来の技術】 オーディオ、画像、およびビデオのような電子化されたメディアが増加すると共にデータ源の確認を容易に確認できるような保証システムが必要になってきている。これは、著作権の施行および情報源の確認のためであることは明らかである。

【0003】 従来の暗号システムを使用した場合、暗号化されたデータに対して正しい鍵所有者だけがアクセスできる。しかし、一度データが暗号化されれば、次の表示あるいは送信の記録を維持できなくなる。従来の暗号化法は、データあるいは情報の出版者あるいは所有者に著作権侵害のデータに対して最小の保護しか提供しないので、無許可の複製あるいはこのようなデータあるいは情報の配布に直面することになる。

【0004】 電子透かしは暗号処理を補完することを企図している。透かしは、データ中に不変的に埋め込まれる、可視あるいは好ましくは不可視な認識符号である。すなわち、透かしはいかなる解読過程のあとでもデータとともに残存している。ここで使用されるデータおよび情報の用語は、オーディオ（スピーチおよび音楽）、画像（イメージ）（写真およびグラフィックス）、ビデオ（ビデオ）（映画あるいは連続画像）およびマルチメディアデータ（上記カテゴリーに属する情報の結合）あるいは、それらの処理されたあるいは圧縮されたものに関するものを指している。これらの用語は、テキストのASCII表記を意図するものではなく、画像として表示されたテキストを指示している。透かしの簡単な例としては、著作権所有者を識別するために画像上に配置する可視“シール”である。しかしながら、透かしは、また、画像の特定の複製を購入した購入者を識別するような付加的な情報も含んでいてもよい。効果的な透かしは、以下の特性を有するべきである：

1. 透かしは、知覚的に不可視であるべきで、その存在が保護される情報に対して邪魔なものであってはならない。
2. 透かしは、意図された目的に対しては情報を無効にすることなく情報から除去されることが困難（好ましくは実質的に不可能）でなければならない。しかしながら、もし部分的にのみ知られて、例えば、画像内の透かしの正確な位置については知られていない場合、例えばノイズを付加することによって、透かしを除去あるいは破壊するというような手段が行われると、透かしが除去あるいは喪失される前に、データの忠実度を著しく低下させ、データを無効にすることが望ましい。
3. 透かしは、複数の人々が共謀して透かし処理されたデータ複写をそれぞれ所有しようとする行為に対して強固なものでなければならない。すなわち、透かしは、透かしを破壊するために設けられた同じデータの複製処理

の組合に対して強固でなければならない。また、その意図的侵害者達が各画像を組み合わせて異なった効果的な透かしを生成することが可能であってはならない。

4. 透かしは、データに共通の信号処理操作を施すことによって、検索可能でなければならない。これらの操作は、例えば、画像のコントラストおよび色、あるいはオーディオの低音および高音に対して、デジタル-アナログ変換、アナログ-デジタル変換、再標本抽出、再量子化（網点処理および再圧縮を含む）、及び共通の信号強調処理等を含んでいる。画像およびオーディオデータにおける透かしは、回転、変換、トリミング、およびスケールリングのような幾何学的画像操作に影響されないようなものでなければならない。

5. 同じ電子透かし方法あるいはアルゴリズムは、検討されている互いに異なったメディアに対して適用可能でなければならない。これは、特に、マルチメディア情報の透かし処理に便利である。さらに、この特徴は、共通のハードウェアを使用して映像及び画像/ビデオ透かし処理の実行に可能にするものである。

6. 透かしの検索の際、所有者が明確に識別されなければならない。さらに、所有者識別の精度は、妨害をうけると低下されるのが望ましい。

【0005】いくつかの電子透かし方法が既に提案されている。L. F. Turnerは“Digital Data Security System”という名称の特許第W089/08915号では、無作為に選択されたオーディオサンプルのうち、“重要でない”ビットを識別符号のビットと置換することによって、識別符号列をデジタルオーディオ信号に挿入する方法が提案されている。もし、変更が聞き取れない場合は、ビットは重要でないみなされる。このようなシステムは、また“A digital watermark”, (Intl. Conf. on Image Processing, vol 2, Pages 86-90, 1994)

という論文において、R. g. Van Schyndel等によって述べられているように、画像のような2次元データに適用可能である。Turnerの方法は簡単に破られてしまうであろう。例えば、もし、アルゴリズムが、単に、ワードの最小下位2ビットにのみ影響を与えることが知られてしまえば、すべてのこのようなビットを任意にはじき出すことが可能になり、それによって存在している識別符号を破壊してしまうことができる。G. Caronniによる“Assuring Ownership rights for Digital Images”, (Proc. Reliable IT Systems, VIS '95, 1995)という論文で、タグ（即ち、小さな幾何学的パターン）を視覚的に感知できない輝度レベルで電子化された画像に付加することを提案している。画像の空間的透かしを隠すという考えが基本的に安全である一方で、この手段は汙染および再電子化によって侵害を可能にする。このような透かしはばやけるほど、侵害が可能となり、幾何学的形状には、符号化すべき情報として、限られたアルファベットしか使用できなくなる。さらにその手段はオーディ

オデータには使用できず、特にトリミングのような共通の幾何学的な歪みには弱い。

【0006】J. Brassil等は、“Electronic Marking and Identification Techniques to Discourage Document Copying”, (Proc. of Infocom 94, pp 1278-1287, 1994)という論文において、テキストが共通である文書画像に適する3つの方法を提案している。電子透かしは以下の方法によって符号化される：(1) 実質的にテキストラインを垂直に移動すること、(2) ワードを水平に移動すること、又は、(3) 個々の文字の垂直な終端線のようなテキストの特徴を変更する。ただ残念なことに、3つの提案全てが、著者によって述べられているように簡単に破られてしまう。さらに、これらの技術はもっぱらテキストを含んでいる画像にのみ限定されている。

【0007】K. Tanakaによる“Embedding Secret Information into a Dithered Multi-level Image”, (IEEE Military Comm. Conf., pp 216-220, 1990)および、K. Mitsui等による“Video-steganography”, (IMA Intellectual Property Proc., vi, pp 187-206, 1994)という論文では、量子化ノイズに似た透かしを埋め込むことによるいくつかの透かし手段について述べられている。彼等の考えは、量子化ノイズが見る人にとっては基本的に知覚できないというコンセプトに基づいている。第1の手段は、所定のデータ列を使用することによって画像に透かしを注入し、予測量子化器において、レベル選択を行っている。データ列は、結果的に生成された透かしが量子化ノイズのように見えるように選択される。この手段の変形例もまた提示されている。それによると、網点処理マトリックスの形状の透かしが、ある方法で画像を網点処理するために使用される。これらの手段にはいくつかの欠点がある。最も重要なものは、それらが、特に量子化のような信号処理、および、トリミングのような幾何学的処理による攻撃を受けやすいことである。さらに、それらは、予測符号化および網点処理と同様に画像の質を低下させる。

【0008】Tanaka等において、ファクシミリデータを透かし処理する手段も提案されている。この手段は、符号化されたファックス画像を生成するために使用されるランレングス符号中における一連のデータを短くしたり長くしたりするものである。この提案は、デジタル-アナログ、およびアナログ-デジタル変換の影響を受けやすい。特に、それぞれのピクセル強度の最下位ビット(LSB)をランダム化することによって結果的なランレングスの符号化が完全に変更される。Tanaka等はまた、“color-scaled picture and video sequences”(カラー画像及び映像列)のための透かし方法を提案している。この方法は、JPEG(画像の8×8のサブブロックのDCT)と同じ信号変換)を使用しており、係数量子化モジュール中に透かしを埋め込んでいる。現

存している変換コードにも使用できるが、この手段は再量子化および歪波の影響を受けやすく、変換係数の最下位ビット中に透かしを符号化することと均等である。

【0009】MacqおよびQuisquaterによる“Cryptology for Digital TV Broadcasting”, (Proc. of the IEEE, 83(6), pp 944-957, 1995) という最近の論文では、暗号およびデジタルテレビに関する一般的な研究の一部として透かし処理した電子画像の問題について簡単に述べている。著者は、画像輪郭の近くに位置しているピクセルの最下位ビットに透かしを挿入する方法を説明している。この方法は、最下位ビットの修正に依存しているので、透かしは簡単に破壊される。さらに、この方法は、輪郭の端部にある画像領域に透かしを挿入できる画像にのみ適用される。

【0010】W. Bender等は“Techniques for Data Hiding”, (Proc. of SPIE, v2420, page 40, July 1995) という論文において、2つの透かし手段について記載している。第1は、“パッチワーク”と呼ばれる統計的な方法である。パッチワークは画像点( $a_i, b_i$ )の $n$ 個のペアをランダムに選択し、 $b_i$ の輝度の減少に対応して、1ユニットの $a_i$ における輝度を増加させている。画像のある統計的特性が真であるという仮定の下に、 $n$ 個の点のペアの差の総計の期待値は、 $2n$ であることが要求される。特に、全ての輝度レベルが等しい、すなわち、強度が均一に分布していることが仮定されている。しかしながら、実際には、これは非常にまれなことである。さらに、この手法は単一ユニット毎に強度レベルにランダムに与えられるジッタに対して強くなく、そして特に幾何学的アフィン変換に対して影響を受けやすい。

【0011】第2の方法は、“texture block coding”と呼ばれるもので、この方法では、画像に見られるランダムなテクスチャパターンの領域が類似のテクスチャを有する画像の領域に複写される。それから、自己相関が各テクスチャ領域を再現するために使用される。この技術に伴う最も重要な問題は、大きなランダムテクスチャ領域を有する画像に対してのみ適用できるということである。例えば、この技術はテキスト画像には使用できない。また、オーディオ用アナログ信号にも直接適用できない。

【0012】透かし処理画像での直接的な研究の他に、関連する領域におけるいくつかの興味ある研究がある。“Digital Signal Encoding and Decoding Apparatus”という名称の米国特許第4,939,515号において、E. H. Adelsonは、アナログテレビ信号にデジタルデータを挿入するために、アナログ信号にデジタル情報を埋め込む技術について述べている。アナログ信号は、送信されるべきバイナリーデジットに基づいて選択される2つの互いに素の集合(例えば、 $\{0, 2, 4, \dots\}$   $\{1, 3, 5\}$ )の1つに量子化される。こ

のようにAdelsonの方法は、情報をデータあるいは変換係数の最下位ビットに符号化する透かし手段と均等である。Adelsonは、この方法はノイズの影響を受けやすいため、 $2 \times 1$ のデジタル化されたアナログ信号のアダマール変換が行われるもう一つの手段を提案している。アダマール変換の微分係数は、逆変換を計算する前に0あるいは1に区分される。これは、アダマール変換における微分係数の最下位ビットに透かしを符号化することに対応している。この処理方法は、ノイズに対して強い回復力を示しているとはいえない。さらに、このような全ての最下位ビット手法では、読者はランダム化することによって透かしを除去できる。

【0013】米国特許第5,010,405号では、拡張解像度テレビ(EDTV)信号内に標準NTSC信号をインターリーブする方法が述べられている。これは、EDTV信号の周波数スペクトム(NTSC信号のそれよりも大きい)を分析し、3つのサブバンド(L、M、Hはそれぞれ低、中および高周波数)に分解することによって達成される。一方、NTSC信号は2つのサブバンドLとMに分解される。Mバンドの係数 $M_k$ はMのレベルに量子化され、EDTV信号の高周波係数 $H_k$ は、 $H_k$ 信号とシステムに存在するノイズの和が量子化レベル間の最小の分離間隔より小さくなるような大きさに選ばれる。さらに、この方法は最下位ビットの修正による。仮定的には、知覚的にはあまり重要でない低周波よりむしろ中範囲が選択されるかもしれない。

【0014】それに対して、本発明で提案されている方法は信号のほとんどの知覚的に重要な成分を修正する。最後に、全部でないが、先行技術のプロトコルの多くが共謀による妨害に対して抵抗力がない。

【0015】最近、オレゴン州のポートランドにあるDigimarc Corporationは、電子的な知的財産を識別するのに使用する署名技術の研究について説明している。彼らの方法は、それぞれのピクセルに対して小さなランダム数を加算したり減算したりするものである。加算あるいは減算は、 $N$ 個のビットのバイナリーマスクをそれぞれのピクセルの最下位ビット(LSB)との比較に基づいて行われている。もし、LSBが対応するマスクビットと等しければ、ランダム数が加算され、そうでなければ減算される。透かしは、まず、オリジナルの画像と透かし処理された画像間の差を計算し、差の正負の符号をピクセル毎に検査し、加算/減算の元の数値に対応するかどうか決定する。Digimarcの技術は、画像スペクトラムの直接的な修正には基づいておらず、知覚的な適合性を使用していない。その技術は強固なようにみえるが、一定の輝度による相殺を受けやすく、且つ、画像に存在する高い局所の相関関係を使用することに基づく妨害の影響を受けやすい。例えば、局部隣接位置内で類似ピクセルの位置をランダムに切り替えると、画像に損失を与えずに、透かしをかなり低下させることになる。

【0016】Koch, RindfreyおよびZhaoによる論文“Copyright Protection for Multimedia Data”では、透かし処理画像の2つの一般的な方法が記載されている。1番目の方法は、画像を $8 \times 8$ のピクセルからなるブロックに分割し、各ブロックの離散余弦変換(DCT)を行う。ブロックの疑似乱数の部分集合が選択され、このようなブロック内で、18個の所定の3重倍周波数から選択された単一の3重倍の周波数が相対強度が1あるいは0の値に符号化するように修正される。18個の可能な3重倍周波数は、 $8 \times 8$ のDCTブロック内の8個の所定の周波数から3つを選択することによって構成される。DCTブロック内で変化させられるべき8つの周波数の選択は、中間周波数が適度な分散レベルを有する、すなわち、相似した大きさを有しているという考えに基づいているようである。この特性は、知覚的に感知できる修正を必要とすることなく3重倍周波数の相対的な強度を変化させるために必要とされる。本発明とは異なり、一組の周波数は、知覚的な重要性、或いは相対的なエネルギーを考慮して選択されたものではない。更に、8つの周波数係数間の分散が小さいので、この技術はノイズあるいは歪みの影響を受けやすい。このことは、Koch等の論文で報告された実験結果によって支持されており、それによると、埋め込まれたラベルは、約50%程度の低い品質因子のJPEGに対して強固であることが報告されている。これに対し、本発明で説明されている方法は、5%低い圧縮品質因子による圧縮にも強固であることが証明されている。

【0017】KochとZhaoは、“Toward Robust and Hidden Image Copyright Labelling”という論文において、3重倍周波数ではなく、2つの周波数対を使用することを提案しており、特に、JPEG圧縮に対する強さについて再度設計を行っている。にもかかわらず、報告書は“より低い品質因子を使用した場合、埋め込まれた符号を信号に重畳するのに必要な変化は感覚的に感知できる傾向を増加させるであろう”と述べている。

【0018】KochとZhaoによって提案されている第2の方法は、ブラックおよびホワイト画像用に設計されたものであり、周波数変換は使用されない。代わりに、ホワイトおよびブラックピクセルの相対周波数が最終値に符号化されるように、選択されたブロックが修正される。両方の透かし処理手順のいずれも、特に複数回になる文書妨害に対して弱い。これに対する保護を与えるために、Zhao及びKochは画像からランダムに64個のピクセルをサンプリングして得られた $8 \times 8$ ピクセルからなる分散されたブロックを使用することを提案した。しかし、結果として生じるDCTは真の画像のそれとは関係がない。従って、このような分散されたブロックが両方ともノイズに対して弱く、画像内に目立った不所望の特性、即ち、アーティファクトを生じさせることもありうる。

【0019】要約すると、先行技術の電子透かし技術は強固なものではなく、透かしは簡単に取り除かれるものである。さらに、多くの先行技術は共通の信号および幾何学的な歪みに耐えられない。

【0020】

【課題を解決するための手段】本発明は、画像、オーディオ信号あるいは映像のシーケンスを分解し、分解部分の知覚的に重要な成分中に独特の識別子を埋め込む透かしシステムを提供することによって従来の方法の制限を克服するものである。

【0021】好ましくは、分解はスペクトラム周波数分解であることである。透かしは、データの知覚的に重要な周波数成分に埋め込まれる。多くの共通信号あるいは幾何学的処理がこれらの成分に影響するので、効果的な透かしが、画像データの知覚的に重要でない領域あるいはその周波数スペクトムに位置することができないのとは異なっている。例えば、画像の高周波数スペクトラム成分に位置する透かしは、低域通過フィルタを行う処理によって画像のわずかな質の低下を伴うだけで、簡単に除去される。このため、この問題の一つは、観察者、すなわち、人間あるいは機械の特徴認識システムによって変更が見られることなくデータ周波数スペクトムの最も重要な領域にどのようにして挿入するかである。スペクトラム成分の変更が小さければ、スペクトラム成分はどのように変更されてもよい。しかしながら、非常に小さい変更でもノイズの存在あるいは意図的な歪みの影響を受けやすい。

【0022】この問題を解決するために、画像データあるいは音声データの周波数領域は通信チャネルと考えられ、これに対して、透かしは、チャネルを通して送信される信号と考えられる。妨害や意図的な歪みは送信された信号から除外されなければならないノイズとして扱われる。妨害は、透かしを施されたデータの有効性を除去、削除あるいは、否定するための意図的な努力である。本発明はデータに透かしを埋め込むことを意図しているが、同じような方法論が、メディアデータとして種々のタイプのメッセージを送信することに適用される。

【0023】データの最下位成分に透かしを符号化するかわりに、本発明は拡散スペクトム通信の概念を適用している。拡散スペクトム通信において狭帯域信号は、単一周波数に存在している信号エネルギーが知覚できない程度に近くなるような広いバンド幅になって送信される。同様な手法で、透かしは、単一のビン(容器)の中のエネルギーが小さくそして知覚できないような多くの周波数のビン(容器)に拡散される。透かしの照合プロセスは、透かしの位置と内容に関して前もって定められているので、多くの弱い信号を単一の信号に、高い信号対雑音比(S/N比)で集合させることが可能である。このような透かしを破壊するためには全ての周波数ビン

に、高振幅のノイズを付加することが必要となる。

## 11

【0024】本発明に従えば、透かしは、分解されたデータの知覚的に最も重要な領域に挿入される。透かし自身は、付加的なランダムなノイズに見えるように考慮されており、画面中くまなく拡散されている。知覚的に重要な成分に透かしを配置することによって、妨害者は画像あるいはデータに悪影響を及ぼすことなく更に、ノイズを付加することが非常に困難となる。実際に透かしは、ノイズのように見え、通信システムに使用されている拡散スペクトラム法と同様にして、画像あるいはデータ中に分散される。

【0025】画像のスペクトラム中に透かしを拡散させることによって、意図的でないあるいは意図的な妨害に対して大きな安全性が保証される。第1に、透かしの位置が明らかではない。第2に、透かしに対する妨害に伴ってオリジナルのデータの品質が著しく低下してしまうというように、周波数領域が選択される。

【0026】画像あるいは音声トラックの周波数領域に配置された透かしは、実際、見たり聞いたりすることは不可能である。このことは、透かし中のエネルギーが単一の周波数係数中において十分に小さい限り、常に当てはまる。さらに、人間の聴覚および視覚システムにおけるマスキング現象の知識を活用することによって、特定の周波数に存在するエネルギーを増加させることが可能である。知覚的なマスキングとは、画像あるいは音声のある領域の情報が、画像あるいは音声のその他の部分の知覚的により卓越した情報によって遮断されるという状態を指している。電子波形を符号化する際、この周波数領域（および、いくつかは時間/ピクセル領域）マスキングの手法が、データを低ビットレートで符号化するために広く活用されている。聴覚的および視覚的システムの両方においてより高い聴覚的あるいは視覚的な高エネルギー、低周波数、及びスペクトラムに対して付加されることは明らかである。さらに、画像および音声のスペクトラル分析の結果は、このようなデータの情報の大部分はしばしば低周波数領域に位置していることが判っている。

【0027】さらに、特に処理されているあるいは圧縮されているデータに対して、人間の知覚的重要性について言及される必要はないが、代わりに、機械の知覚的重要性、例えば、機械における特徴認識について触れておく。

【0028】これらの要求を満たすために、ゼロ平均および均一分散性を有する通常分布でランダムに生成された、例えば1000という大量の数の透かしが提案されている。画像の数種類の独立して透かし処理された画像を共謀して複写するような妨害に対しては余り強いものでなくなるので、バイナリーの透かしは選択されない。しかし、一般的には、透かしは、均一の分布を含む、決定的および/又はランダムな構造のうち、任意の構造を有してよい。提案されている透かしの長さは可変で

## 12

あり、データの特性に合うように調整され得る。例えば、より長い透かしは、スペクトラム係数の大幅な修正に対して特に敏感なため、個々の成分に弱いスケールリング因子を必要とする画像に使用される。

【0029】透かしは、画像スペクトラムの成分中に配置される。これらの成分は、最も妨害に対して弱い、および/又は最も知覚的に重要である成分を分析した結果に基づいて選択される。これは、透かしが、共通信号および幾何学的歪みの後でさえ画像に残ることを保証している。これらのスペクトラム成分の修正の結果、透かしが破壊されるよりずっと以前に著しい画像の質の低下が生じる。透かしを挿入するためには、同一の係数を変更することが必要であることは勿論である。しかしながら、それぞれの修正は極端に小さくすることができ、拡散スペクトラム通信と類似した方法で強固な狭帯域の透かしが、より広い画像（チャネル）スペクトラム中に分散させられる。概念的に言えば、透かしは、著作権所有者しか知られないような位置にある非常に小さい信号の全てを加え合わせ、透かしを高いS/N比で単一の信号に集合させることによって検出できる。透かしの位置は著作権所有者にしか知らされていないので、妨害者は、透かしを確実に除去するためには、それぞれのスペクトラム係数により多くのノイズエネルギーを付加しなければならない。しかし、この処理は画像を破壊してしまうことになる。

【0030】好ましくは、DCT（離散余弦変換）（DC成分を除く）の最も大きな係数が所定数だけ使用される。しかし、DCTを選択することはアルゴリズムあるいはその他のスペクトラム変換に対して決定的なことではなく、ウェーブレット形式に分解してもよい。事実、計算上における見込みから言えば、DCTよりもむしろFFTを使用することが、好ましい。

【0031】本発明は、添付図面に関する以下の説明から、より明確に理解されるであろう。

【0032】

【発明の実施の形態】本発明の利点をより理解するために、周波数スペクトラムに基づく透かしシステムの好ましい実施の形態を説明する。画像（あるいは音声）データが複写過程で受ける処理段階を検査し、このような処理段階においてデータ上に及ぼされる効果を考えることは有益である。図1を参照すると、透かし処理された画像あるいは音声データ10は送信12され、代表的な歪み処理あるいは、意図的な改ざん処理（以下タンパリング処理と呼ぶ）14を受ける。このような歪みあるいはタンパリング処理は損失のある圧縮処理16、幾何学的変形処理18、信号処理20およびD/AおよびA/D変換処理22を含む。歪みあるいはタンパリング処理を受けた後、改ざんされ、且つ透かし処理された画像および音声データ24は送信26される。“送信”の過程は、供給源あるいはチャネル符号を使用すること、およ

13

び／又はデータに対して暗号化技術処理を施すことを指している。大部分の送信段階において情報に損失は生じないが、大部分の圧縮手段（例えばJ P E G、M P E G等）では、回復できないデータの損失によってデータの品質を潜在的に低下させている。一般的に、透かしの方法は、送信あるいは圧縮アルゴリズムによって導かれる歪みに対して回復可能でなければならない。

【0033】損失圧縮処理16は、通常、画像あるいは音声データの知覚的に関係ない成分を除去する操作である。損失圧縮処理の際に、透かしを保存するために、透かしはデータの知覚的に重要な領域に配置されている。このタイプのほとんどの処理は周波数領域で行われる。データ損失は通常、高周波数成分で発生する。このように、透かしは、損失圧縮による悪影響を最小限にするために画像（あるいは音声）データスペクトラムの重要な周波数成分位置に配置されなければならない。

【0034】受信後、画像は、幾何学的変形あるいは信号の変形として広く類別される多くの共通の変換を受ける。幾何学的変形18は画像および映像データに特定され、回転、変換、スケーリング、およびトリミングといった操作を含む。原透かしと変形された透かしの間の4つあるいは9つの対応する点の最小値を手動的に決定することによって、2あるいは3次元のアフィン変換を除去することが可能である。しかしながら、画像のアフィンスケーリング（萎縮）は画像の高周波数スペクトラム領域におけるデータ損失を招く。画像の部分のトリミング、あるいは切除および除去はまた、回復できないデータの損失を招く。トリミングは空間的な透かしに対して大きな影響を与えるが、周波数上の処理に対してはさほど影響しない。

【0035】共通信号の変形は、網点処理、および再圧縮を含むデジタルーアナログ、アナログーデジタル変換処理22、再標本抽出、再量子化処理、および画像コントラストおよび／又は色に対する共通信号の増幅処理と、オーディオ周波数等化処理を含む。これらの変形の多くは非線形で、空間あるいは周波数上の処理による方法のいずれにおいても効果を分析することは難しい。しかしながら、原画像が知られているという事実のために、少なくともおよそ近似的には多くの信号変換が出来ないようにする。例えば、棒グラフ等化、共通の非線形コントラスト強調手法は、棒グラフ示像あるいは動的棒グラフ変形技術によって実質的には除去される。

【0036】最終的に、複写された画像はデジタル形式のままでは残らない。代わりに、印刷されるか、あるいはアナログ記録（アナログオーディオあるいはビデオテープ）される。これらの再生は、透かしの手法が強固でなければならない画像データに対して付加的な低下をもたらす。

【0037】タンバリング（すなわち妨害）処理は、透かしを除去あるいは認識出来ないようにそれを改ざんす

14

るという意図的な試みのことである。透かしは、不注意による変造に対して耐性を示すだけでは不十分である。悪意のある人達による意図的な操作から免れなければならない。これらの操作は変形の結合を含んでおり、共謀および偽造による妨害をも含んでいる。

【0038】図2は、周波数領域における画像に透かしを挿入する好ましいシステムを示している。デジタル形式のデータ、あるいは前もって周知の方法で電子化されている写真、絵画のようなその他の形式のデータである画像データ $X(i, j)$ は、フーリエ変換のような周波数変換30を受ける。透かし（ウォーターマーク）信号 $W(k)$ は、以下に説明されている技術を適用して変換された画像データ32の周波数スペクトラム成分に挿入される。透かし信号を含んでいる周波数スペクトラム画像データは、逆周波数変換34を受け、透かし処理された画像データ（数式1で示す）となり、このデータは、デジタル形式のままであっても、周知の方法でアナログ印刷されてもよい。

【0039】

【数1】

$$\hat{X}(i, j)$$

画像データに周波数変換30を適用した後、知覚的なマスクが計算され、このマスクは知覚的忠実度に過度に影響を与えることなく、透かしを支持できる周波数スペクトラム中の主要な領域を強調する。これは、前にも述べたように、スペクトラム中における各周波数の知覚的重要性を知ることによって、あるいは単に、それらのエネルギーに基づき周波数を分類することによって実行できる。後者の方法が、以下に説明される実験で使用された。

【0040】一般的に、共通信号の変形の際、ほとんど影響を受けることなく、且つ重要な修正によって画像の忠実度が破壊されるようなスペクトラム領域で観察者によって認識されるような、しかも画像品質に最も重要であるスペクトラムの領域に透かしを配置することが望ましい。実際、これらの領域は、共通信号の変形を画像に適用し、どの周波数が最も影響を受けるかを検査し、画像における重要な変更が知覚されるまでに、いかに多くの成分が修正されるかを認識する精神物理学によって実験的に識別される。

【0041】透かし信号は、それから、タンバリング処理の際にデータ内に可視（可聴）の欠陥を生成させるような方法でこれらの主要領域に挿入される。上述した透かしの要求、および複写に共通している変形は、電子的透かしの設計に制限を加えることになる。

【0042】透かし方法をよりよく理解するために、図3(a)および図3(b)について説明がなされるが、ここでは、それぞれの文書Dから一連の数列 $X=$

$x_1, \dots, x_n$ が抽出40され、この数列と透かしW

15

$=w_1, \dots, w_n$  とが結合42され、調整された数列  $X' = x'_1, \dots, x'_n$  が作り出される。この調整された数列は、透かし文書  $D'$  を得るために、文書の値  $X$  の位置に戻される(44)。文書  $D'$  の妨害、あるいはその他の変形が加わって文書  $D^*$  を作成する。オリジナル文書  $D$  と文書  $D^*$  を有しているのを、改変されたかもしれない透かし  $W^*$  が抽出46され、そして統計的な分析50のために透かし  $W$  と比較48される。 $W^*$  の値は、まず、一群の値  $X^* = x_1^*, \dots, x_n^*$  を  $D^* \cdot$

$$x'_1 = x_1 + \alpha w_1$$

$$x'_1 = x_1 (1 + \alpha w_1)$$

$$x'_1 = x_1 (e^{\alpha w_1})$$

数式2の(1)式は可逆的であり、数式2の(2)式および(3)式は  $x_i \neq 0$  のとき可逆的である。それゆえ、 $X^*$  が与えられているとき、 $X$  および  $X^*$  から  $W^*$  を抽出するのに必要な逆関数を算出することが可能である。

【0045】数式2の(1)式は値  $x_i$  が広範囲に亘って変化するとき、好ましい式とは言えない。例えば、もし  $x_i = 10^6$  のとき、100を加えても、透かしを確立するには不十分であるが、もし  $x_i = 10$  のとき、100を加えることは、容認しがたい程、値を歪ませる。数式2の(2)式および(3)式を使用する挿入方法は、広い範囲の値  $x_i$  に対して強固である。また、数式2の(2)式および(3)式は  $\alpha w_i$  が小さいとき、同様な結果を与えることが認められる。さらに、 $x_i$  が正の時、オリジナルの値の自然対数を使用される場合、数式2の(3)式は  $\ln(x_i) = \ln(x_i) + \alpha w_i$  と等しく、数式2の(1)式の応用として考えられる。例えば、もし  $|w_i| \leq 1$  で、且つ、 $\alpha = 0.01$  であれば、数式2の(2)式を使用した場合、スペクトラム係数は、わずか1%以下しか変化しないことが保証されている。

【0046】ある応用に対して、単一のスケーリングパラメータ  $\alpha$  を、 $x_i$  の全ての値に使用することは最適とは言えない。それゆえ、多数のスケーリングパラメータ  $\alpha_1, \dots, \alpha_n$  が、 $x_i = x_i (1 + \alpha_i w_i)$  のような改訂された数式2の(1)乃至(3)式と共に使用される。 $\alpha_i$  の値は、文書品質を知覚的に変化されるために、 $x_i$  がどれ位変更されなければならないのかを表す相対的な尺度として役立つ。 $\alpha_i$  が大きな値を取る場合、知覚的に文書の品質を低下させずに多くの量の  $x_i$  を変更することが可能であることを意味している。

【0047】多数のスケーリング値を選択する方法は、ある一般的な仮定に基づいている。例えば、数式2の(2)式は、一般化された数式2の(1)式において、※50

16

\* (Dに関する情報を使用) から抽出し、値  $X^*$  および値  $X$  から  $W^*$  を生成することによって値  $W^*$  を抽出する。

【0043】値  $X$  と透かし値  $W$  を段階42で結合するとき、スケーリングパラメータ  $\alpha$  が特定される。スケーリングパラメータ  $\alpha$  は、値  $W$  が値  $X$  を変更する程度を決定する。 $X'$  を算出する3つの好ましい演算式を数式2に示す：

【0044】

【数2】

(1)

(2)

(3)

※  $\alpha_i = \alpha x_i$  としたとき ( $x'_i = x_i + \alpha_i x_i$ ) の特殊なケースである。すなわち、数式2の(2)式からも明かなように、透かし成分  $w_i$  と、周波数成分  $x_i$  とが乗算されており、 $x_i$  が大きくなると、結果的に  $x_i \alpha w_i$  の加算は大きくなる。しかし、 $x_i$  は  $x_i \alpha w_i$  よりかなり大きいので、問題は生じない。

【0048】一般的に、異なった  $\alpha_i$  の値に対する画像の感度は知られていない。経験的に感度を推定する方法は、原画像上の数多くの妨害によって生じる歪みを決定することである。例えば、 $D$  から品質の低下した画像  $D^*$  を算出し、対応する値  $x_1^*, \dots, x_n^*$  を抽出し、そして偏差  $|x_1^* - x_1|$  に比例する  $\alpha_i$  を選択することが可能である。より強固さが要求される場合、歪みのその他の形式を試行して、 $|x_1^* - x_1|$  の平均値に  $\alpha_i$  を比例させることが可能である。平均的な歪を使用する代わりに、中間あるいは最大限の偏差を使用することが可能である。

【0049】他方、値の感度に関する一般的な仮定と実験的手法と結合することも可能である。例えば、 $x_i \geq x_j$  である限り  $\alpha_i > \alpha_j$  である。このことは、数式3に従って  $\alpha_i$  を設定することによって、実験的手法と結合することが可能である：

【0050】

【数3】

$$\alpha_i \sim \max \{ v_j^* - v_j \mid \{ j \mid v_j \leq v_i \} \}$$

より巧みな手法は、突発的に発生する外部から妨害に対して強固にするために、単調性に関する制限を弱めることである。

【0051】透かしの長さ  $n$  は、透かしが画像データの関連のある成分中に拡散している度合いを定めている。透かしのサイズが大きくなると、変更されるスペクトラム成分の数も増え、ノイズに対して同じ強さを維持する

17

という条件の下では、各成分の大きさは長さに応じて小さくなる。ここで、 $x_i' = x_i + \alpha w_i$  の形式をとる透かしと、 $x_i' = x_i + r_i$  の白色雑音による妨害を考慮してみる。この場合、 $r_i$  には、標準偏差  $\sigma$  を伴う独立した正規分布に従って選択されている。 $\alpha$  が  $\sigma / \sqrt{n}$  に比例する時、透かしを回復することが可能である。すなわち、成分の数を4倍にすれば各成分に配置された透かしの大きさを半分にすることができる。偏差の2乗の総和は、本質的に変化しない。

【0052】一般的に透かしは、任意の実数列  $W = w_1, \dots, w_n$  を有している。実際には、それぞれの\*

$$\text{sim}(W, W^*) = \frac{W^* \cdot W}{\sqrt{W^* \cdot W^*}}$$

類似性  $(W, W^*)$  が大きな値を取る場合、この値は、以下の分析において重要である。文書  $D^*$  の著者は  $W$  にはアクセスしないと仮定する（販売人を介するか、あるいは透かし処理された文書を介するかのいずれかによってアクセスする）。どのような  $W^*$  の値が得られても、 $w_i$  に関する条件付分布が  $N(0, 1)$  に従って独立の20に分布する。この場合、数式5が成り立つ。

【0055】

【数5】

$$N(0, \sum_{i=1}^n x_i^2) = N(0, W^* \cdot W^*)$$

このように、類似性  $(W, W^*)$  は  $N(0, 1)$  に従って分布する。それから、正規分布に対する標準の有意性検定が適用される。例えば、 $D^*$  が  $W$  とは関係なく選択されていれば、類似性  $(W, W^*) > 5$  となることはない。類似性  $(W, W^*)$  の幾分高めの値は、たくさんの透かしが記録されている時に必要となることは注意すべきである。上記の分析は  $W^*$  から  $W$  の独立性を要求しており、 $W^*$  それ自身の特定の特性には依存していない。この事実は、 $W^*$  を前処理する際に、さらに柔軟性を与える。

【0056】透かし  $W^*$  の抽出は、透かしを抽出する能力を潜在的に高めるいくつかの方法で行われる。例えば、 $W^*$  の平均値が  $E_i(W^*)$  で示され、網点処理の効果のために実質的に0ではない事例について、画像に関する実験が行われた。この場合に生じる不所望な特性（アーティファクト）は抽出処理の一部において簡単に除去されるが、抽出された透かしを後処理するきっかけを提供する。その結果、簡単な変換、 $w_i^* \leftarrow w_i - E_i(W^*)$  によって優れた類似性  $(W, W^*)$  の値が得られることが分かった。低下した値  $W^* \cdot W^*$  から、改善された性能が得られ、 $W^* \cdot W$  の値はほんのわずかし

か影響を受けなかった。  
【0057】実験において、 $w_i^*$  が、いくつかの  $i$  の値に対してひどく歪められることがしばしば観測され ※50

18

\* 値  $w_i$  は正規分布  $N(0, 1)$  あるいは、 $\{1, -1\}$  又は  $\{0, 1\}$  の一様分布から独立して選択される。 $N(0, 1)$  は平均値  $\mu$  と分散量  $\sigma^2$  を持つ場合、 $N(\mu, \sigma^2)$  によって一般化される。

【0053】抽出されたマーク  $W^*$  がオリジナルの透かし  $W$  と一致することは、ほとんどない。透かし処理された文書を送信のために再量子化したときでさえ、 $W^*$  は  $W$  から変位する。 $W$  および  $W^*$  の類似性の好ましい基準は、数式4で与えられる。

【0054】

【数4】

(4)

※た。一つの後処理選択肢として、このような値を0に設定して、単に無視してもよい。すなわち、数式6としてよい。

【0058】

【数6】

$$w_i^* \leftarrow \begin{cases} w_i^* & \text{if } |w_i^*| > \text{tolerance} \\ 0 & \text{otherwise} \end{cases}$$

このような変換の目的は、 $W^* \cdot W^*$  を小さくすることである。この手法のあまり珍しくないやり方として、 $-1, 0$ 、あるいは1のいずれかであるべき  $W^*$  の値を数式7によって正規化する方法がある。

【0059】

【数7】

$$w_i^* \leftarrow \text{sign}(w_i^* - E_i(W^*))$$

この変換は、その結果の統計的な重要性において劇的な効果がある。その他の強い統計的な技術もまた、外部からの妨害を抑圧するために使用できる。

【0060】原則として、いかなる周波数領域変換でも使用できる。以下に述べる手段ではフーリエ領域における処理が使用されているが、ウェーブレットの手法も変形例として使用可能である。変換の周波数領域を選択するために、問題となる知覚的なシステムのモデルを使用することが可能である。

【0061】信号が、ウェーブレット変換あるいは多変分解変換によってサブバンドに分割される場合、周波数分析はウェーブレット変換またはサブバンド変換によって行われる。各サブバンドは、均一に間隔をおいて配置される必要はない。各サブバンドは、信号の周波数範囲のサブ領域に対応する周波数領域をあらわすものと考えられて良い。この場合、透かしはそのサブ領域に挿入される。

【0062】オーディオデータでは、“ウィンドウ”を信号データに沿って移動させ、ウィンドウ内サンプルについて周波数変換(DCT, FFT等)が行われる。こ

の過程により、時間的に変かしている実際の信号から意味のある情報を得ることができる。

【0063】周波数領域の各係数は知覚的な受容力を有するものと仮定される。すなわち、データの知覚的忠実度に対して何等の悪影響なく（あるいは最小限の悪影響の下で）付加的な情報を挿入できる。

【0064】 $N \times N$ 画像に長さ $L$ の透かしを配置するために、画像の $N \times N \cdot FFT$ （あるいは $DCT$ ）が実行され、 $DC$ 成分を除いて、変換マトリックス中の $L$ 個の最も大きい係数に、透かしが配置される。さらに一般化すると、ランダムに選ばれた $L$ 個の係数は $M$ から選択され、 $M$ は変換の最も知覚的に重要な係数であって、 $M \geq L$ の関係がある。大部分の画像では、これらの係数は低い周波数に対応するものである。透かしをこれらの位置に配置するのは、これらの周波数に関する意味のあるタンパリング処理によって、透かしが破壊される前に、画像の忠実度および知覚される品質が損なわれるからである。

【0065】 $FFT$ は、 $DCT$ に知覚的には類似した結果を与える。これは、変換符号化の場合とは異なっている。変換符号化の場合にはそのスペクトラム特性のために $DCT$ が $FFT$ により望ましい。 $DCT$ では、 $FFT$ より、高周波数情報において少なくなる傾向があり、画像情報のほとんどが低周波数領域に配置される。このことは、除去されるべきデータがある場合には好ましい。透かしの場合において、画像データは保存され、除去されるものは何もない。このように、 $FFT$ は $DCT$ 同様に適用でき、むしろ、計算の容易性の面ではより好ましい。

【0066】実験において、視覚的に知覚出来ない透かしが画像に意図的に配置された。次に、ランダムに100個の透かしが生成され、そのうちの一つだけが正しい透かしに対応付けられた。このような透かしが、上述の透かし検出器に適用された。その結果、図4に示すように、正しい透かしに対応して非常に強い正の応答を示している。このことは、この方法が、正の誤った応答及び負の誤った応答の割合を非常に少なくできることを示唆している。

【0067】もう一つの試験において、透かし処理された画像にオリジナル画像の大きさの半分に定められた透かしを再生するために、低域通過空間フィルタ動作を使用して画像をサブサンプリングし、画像をオリジナルの大きさに戻された。この場合、微細な部分での損失が生じた。透かし検出器の応答は、透かしの存在しない場合におけるランダム確率レベルより十分高く、このことは、透かしが幾何学的な歪みに対して強固であることを示唆している。オリジナルデータの75%が縮尺された画像から失われても、同じ結果が得られた。

【0068】さらなる実験において、10%の品質および0%の平滑度を有するパラメータで画像をJPEGに

従って符号化した。透かし検出器の結果は、この方法が共通の符号化された平滑度を歪みに対して強固であることを示している。5%の品質および0%の平滑度を有するパラメータで画像を使用したときでさえ、ランダム確立によって達成可能な値より、良好な結果が得られた。

【0069】画像の網点処理を使用する実験において、透かし検出器の応答はその方法が共通符号化の歪みに対して強固であることを示している。さらに、抽出された透かしから非ゼロ平均値を除去することによってより信頼できる検出を行うことができる。

【0070】もう一つの実験では、画像の中央の4分の1のみを残して画像が切り取られた。透かしを切り取られた画像から抽出するために、画像の失われた部分が、オリジナルの透かし処理されていない画像の部分と置き換えられた。透かし検出器は、ランダムの場合より多くの応答により、透かしを取り出すことができた。非ゼロ平均値が除去され、透かしの要素が正しい透かしとの比較の前に2倍に値化されると、検出器の応答は改善された。この結果は、データの75%が画像から除去されたときでも達成できた。

【0071】もう一つの実験において、画像は、印刷され、写真複写され、300dpi・Umax・PS-2400x・スキャナーを使用して走査され、256x256ピクセルの大きさに戻された。明らかに、この最終段階の画像は、各処理で加えられた異なったレベルの歪みの影響を受けていた。高い周波数におけるパターンノイズは特に目立っていた。非ゼロ平均値が除去され、透かしの成分の符号だけが使用されたとき、透かし検出器の応答はランダム確率レベル以上にまで改善された。

【0072】もう一つの実験において、画像は連続的に5回の透かし処理された。すなわち、オリジナル画像が透かし処理され、透かし処理された画像が、更に透かし処理された。この処理は、この処理が繰り返されれば、重大な画像の品質低下を招くような妨害のもう一つの形態と考えられて良い。図5は、画像に5つの透かしを含ませた場合にランダムに発生された1000個の透かしに対する透かし検出器の応答を示している。グラフ中の5つの主要なスパイクは5つの透かしの存在を表しており、5つの主要なスパイクは、連続的な透かし処理に何等妨害をも引き起こさないことを示している。

【0073】連続的に透かし処理ができることは、透かしが連続的に各複写に付加されるのであれば、文書の経歴あるいは経路を決定できることを意味している。

【0074】多重化された透かし画像の変形において、別々に施された5つの透かし画像が、平均化されることにより、簡単な共謀による妨害をシュミレートしている。図6は、オリジナル画像に5つの透かしを含ませた場合、1000個のランダムに発生された透かしに対する透かし検出器の応答を示している。その結果は、平均化することになる簡単な共謀では、本発明の透かし処理

21

システムを妨害するのには有効ではないことを示している。

【0075】上記の実験の結果は、数種類の共通の幾何学的および信号処理手段によって大きく品質の低下した画像から、信頼できる透かしの複写を抽出することを示している。これらの手段は、拡大（低域通過濾波処理）、トリミング処理、損失JPEGの暗号化、網点処理、写真複写、および連続した再走査処理を含んでいる。

【0076】これらの実験は、実際には、画像を使用し10て行われたが、同様な結果はテキスト画像、オーディオデータおよびビデオデータにおいても得られる。この場合、データの時間的な変化を考慮する必要があることは勿論である。

【0077】上記した透かし処理システムは電子システムで実現されている。本発明の基本的な原則はデータの分光周波数成分の中に透かしを含んでいることであるので、透かし処理は、例えば、図7に示されている光学的システムを使用しても達成できる。

【0078】図7において、画像40のような透かし処理されるべきデータが、例えば、フーリエ変換レンズのような空間的変換レンズ42を介して通過する。レンズの出力は画像を空間的に変換したものとなる。同時に、透かし画像44は第2の変換レンズ46を通過し、レンズの出力は透かし画像44を空間的に転送したものとなる。レンズ42からの空間的変換結果およびレンズ46からの空間的変換結果は光学結合器48で結合される。光学結合器48の出力は、逆空間変換レンズ50を

22

通過し、その出力には透かし画像52が存在している。その結果は、単一性で実質上不可視な透かし処理された画像である。同様な結果が、上述の方法により、映像あるいはマルチメディアデータをレンズを通過させることによって達成可能である。

【0079】データの拡散スペクトラム透かし処理およびその変形およびその修正を説明してきたが、当業者には、ここに付加した請求項の範囲によってのみ制限される本発明の広い原則および精神から逸脱することなく、さらなる変形および修正が可能であることは明らかである。

【図面の簡単な説明】

【図1】データが受ける基本的な共通の処理操作を示す概略図である。

【図2】透かしを画像に埋め込む好ましいシステムを示す概略図である。

【図3】(a) および (b) は、透かしの符号化および複号化動作を示すフローチャートである。

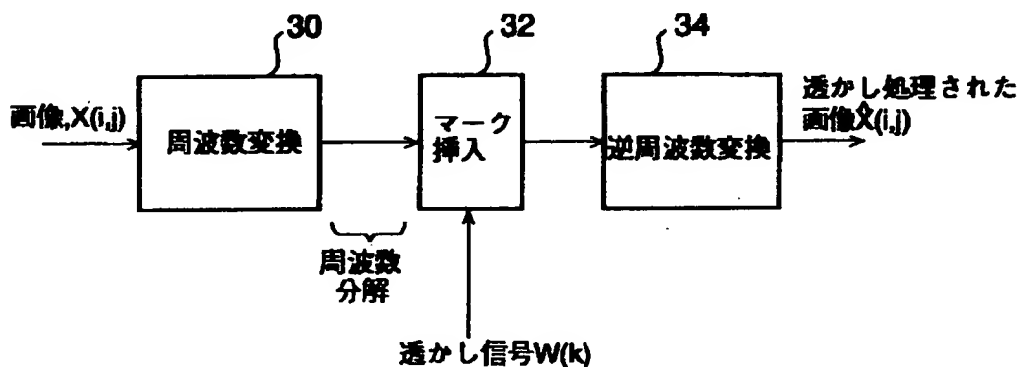
【図4】ランダムな透かしに対する透かし検出器の応答のグラフである。

【図5】連続的に5回透かし処理された画像に対するランダムな透かしを検出する透かし検出器の応答を示すグラフである。

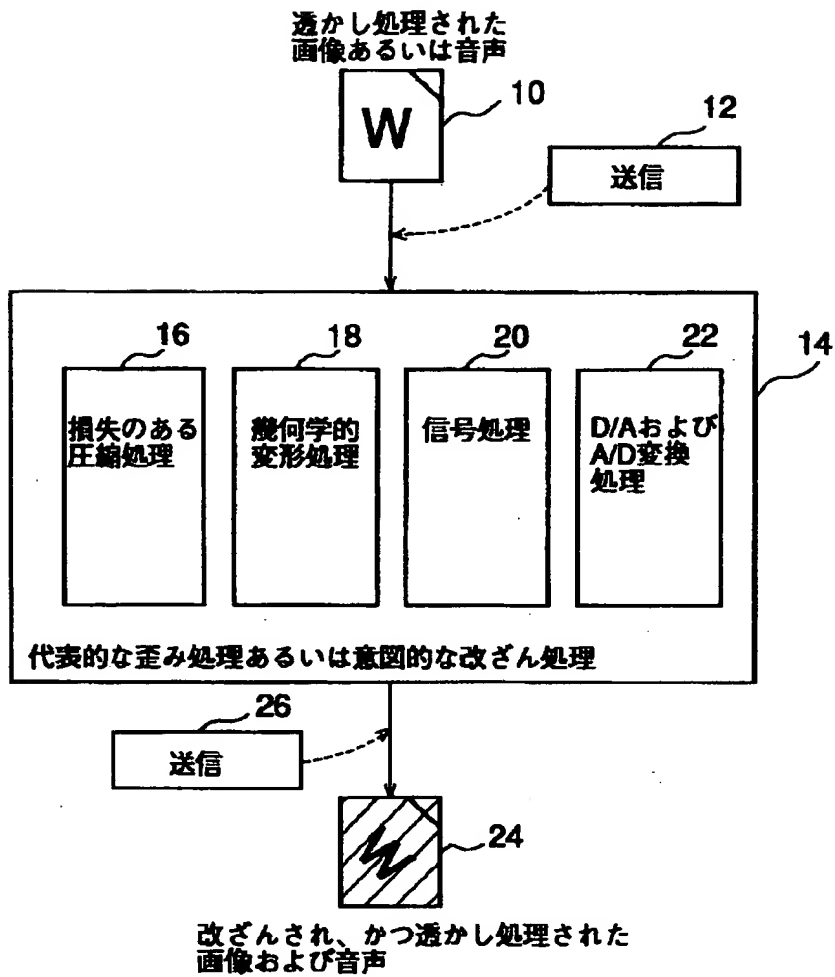
【図6】5つの画像がそれぞれ異なった透かしを有していて、互いに平均されているようなランダムな透かしに対する透かし検出器の応答のグラフである。

【図7】本発明の光学的実施例の概略図である。

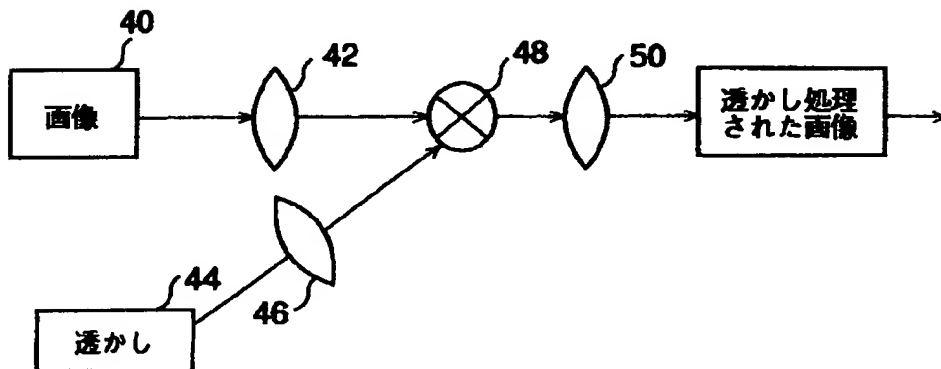
【図2】



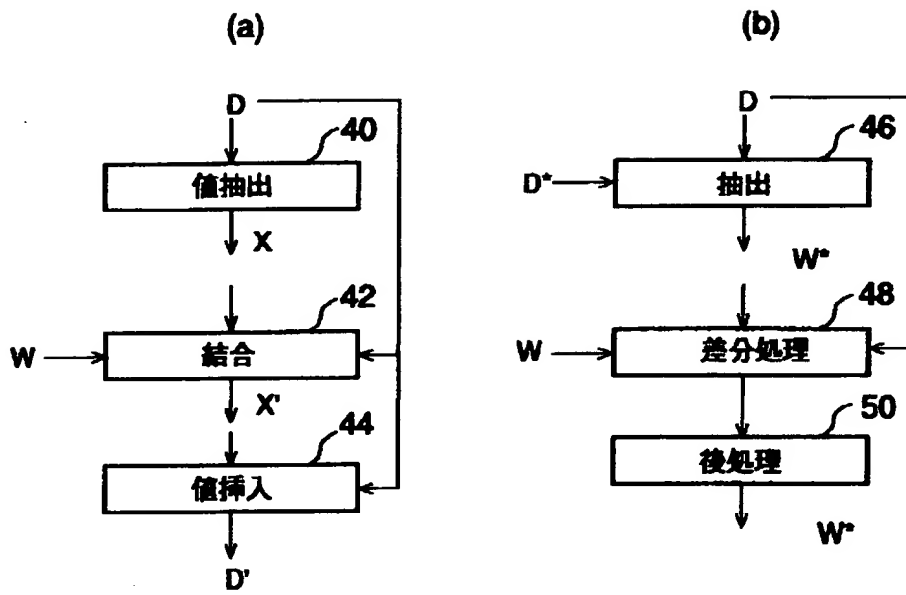
【図1】



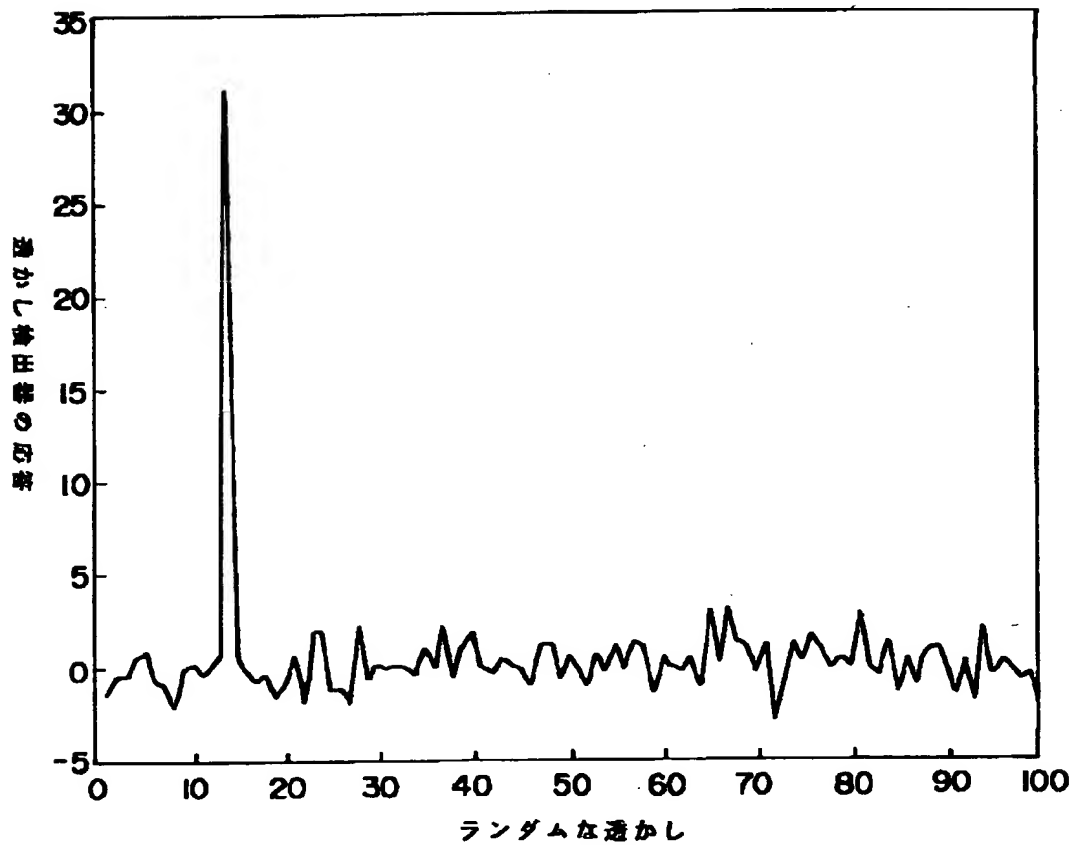
【図7】



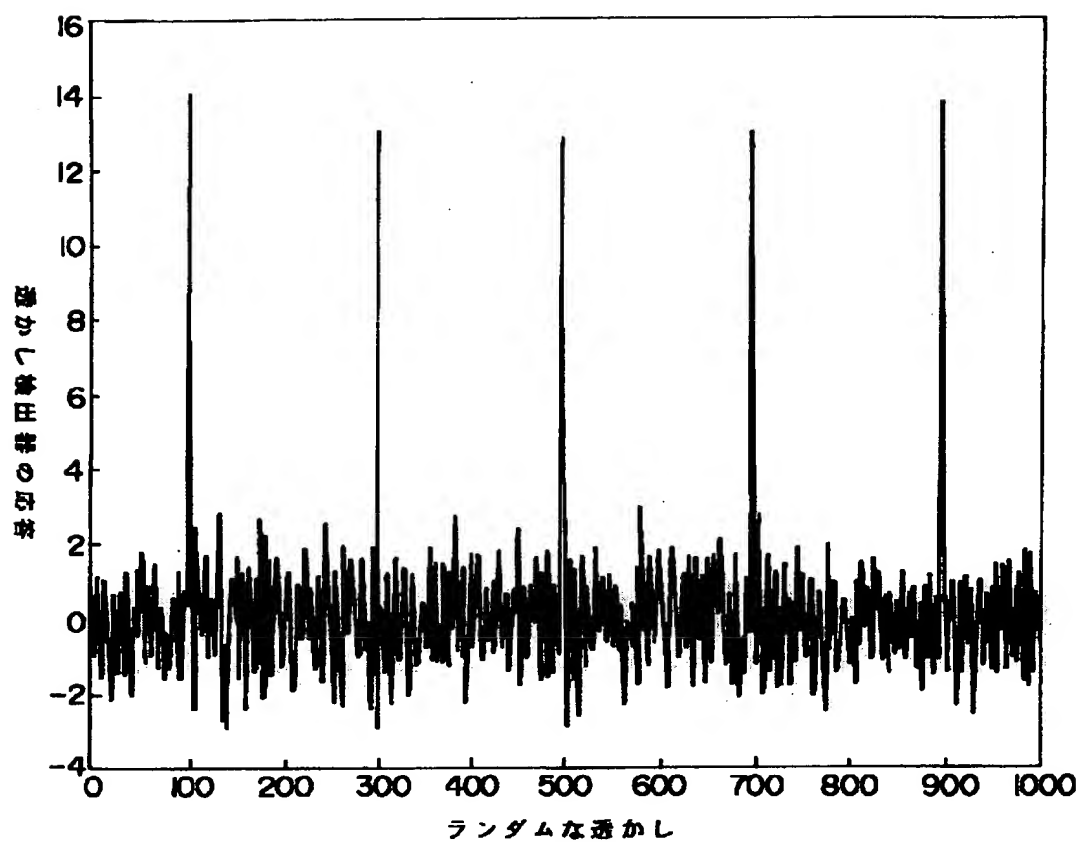
【図3】



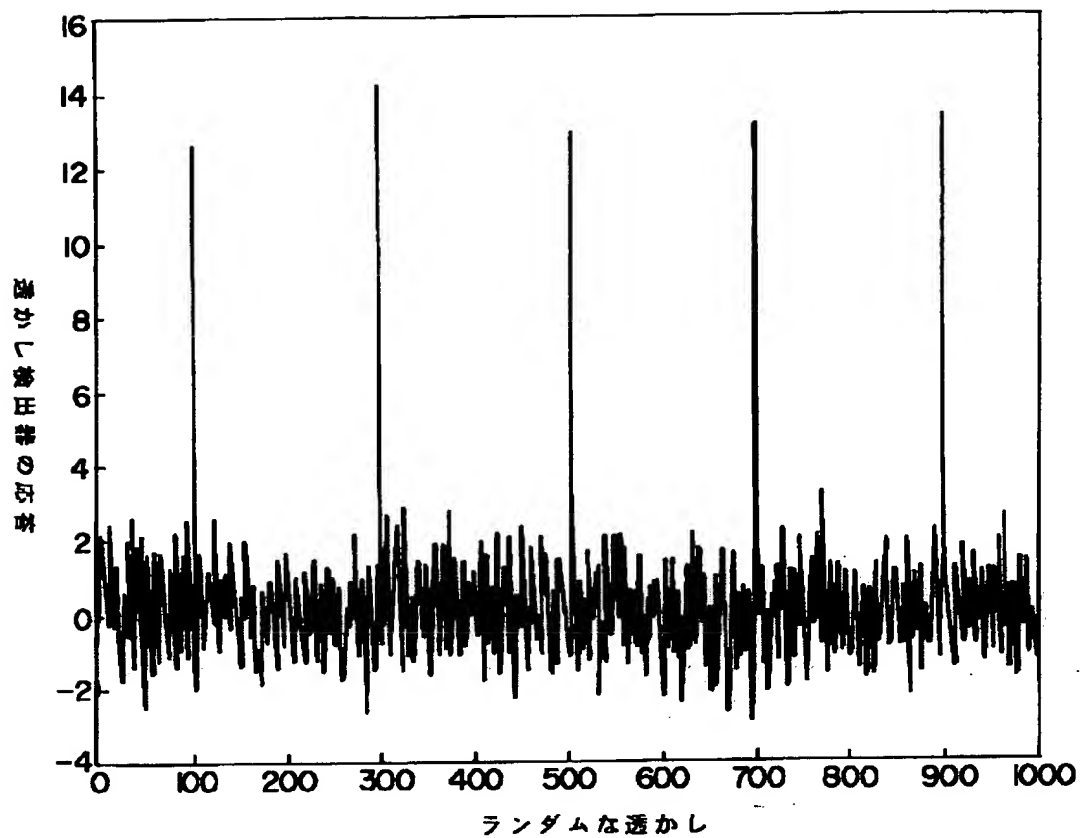
【図4】



【図5】



【図6】



フロントページの続き

(72)発明者 タラル シャムーン  
アメリカ合衆国、ニュージャージー  
08540, プリンストン, ホッジ ロード  
142エー

## 【外国語明細書】

## 1. Title of Invention

Method of inserting a Watermark into Data

## 2. Claims

1. A method of inserting a watermark into data comprising the steps of:  
obtaining a decomposition of data to be watermarked;  
  
inserting a watermark into the perceptually significant components of the  
decomposition of data; and  
  
applying an inverse transform to the decomposition of data with the watermark for  
generating watermarked data.
2. A method of inserting a watermark into data as set forth in claim 1, where said data  
comprises image data.
3. A method of inserting a watermark into data as set forth in claim 1, where said data  
comprises video data.
4. A method of inserting a watermark into data as set forth in claim 1, where said data  
comprises audio data.
5. A method of inserting a watermark into data as set forth in claim 1, where said data  
comprises multimedia data.

6. A method of inserting a watermark into data as set forth in claim 1, where said inserting a watermark inserts watermark values so that addition of additional signal into a perceptually significant component affects the perceived quality of the data.
7. A method of inserting a watermark into data as set forth in claim 1, said obtaining a decomposition of data being obtaining a spectral decomposition of data.
8. A method of inserting a watermark into data as set forth in claim 7, where said data comprises image data.
9. A method of inserting a watermark into data as set forth in claim 7, where said data comprises video data.
10. A method of inserting a watermark into data as set forth in claim 7, where said data comprises audio data.
11. A method of inserting a watermark into data as set forth in claim 7, where said data comprises multimedia data.
12. A method of inserting a watermark into data as set forth in claim 7, where said obtaining a spectral decomposition of data is selected from the group consisting of Fourier transformation, discrete cosine transformation, Hadamard transformation, and wavelet, multi-resolution, sub-band method.
13. A method of inserting a watermark into data as set forth in claim 12, where said inserting a watermark inserts watermark values so that addition of additional signal into a perceptually significant component affects the perceived quality of the data.

14. A method of inserting a watermark into data as set forth in claim 7, further comprising:  
  
comparing data with watermarked data for obtaining extracted data values;  
  
comparing extracted data values with watermark values and data for obtaining difference values; and  
  
analyzing difference values to determine the watermark in the watermarked data.
15. The method of inserting a watermark into data as set forth in claim 14, where watermark values include associated scaling parameters.
16. A method of inserting a watermark into data as set forth in claim 15, where scaling parameters are selected such that adding additional watermark value affects the perceived quality of the data.
17. A method of inserting a watermark into data as set forth in claim 14, where the watermark values are chosen according to a normal distribution.
18. A method of inserting a watermark into data comprising the steps of:  
  
extracting values of perceptually significant components of a spectral decomposition of data;  
  
combining watermark values with the extracted values to create adjusted values; and  
  
inserting the adjusted values into the data in place of the extracted values to produce watermarked data.

19. The method of inserting a watermark into data as set forth in claim 18, where watermark values include associated scaling parameters.
20. A method of inserting a watermark into data as set forth in claim 19, where scaling parameters are selected such that adding additional watermark value affects the perceived quality of the data.
21. A method of inserting a watermark into data as set forth in claim 18, where the watermark values are chosen according to a random distribution.
22. A method of inserting a watermark into data as set forth in claim 18, further comprising:  
  
comparing data with watermarked data for obtaining extracted data values;  
  
comparing extracted data values with watermark values and data for obtaining difference values; and  
  
analyzing difference values to determine the watermark in the watermarked data.
23. The method of inserting a watermark into data as set forth in claim 22, where watermark values include associated scaling parameters.
24. A method of inserting a watermark into data as set forth in claim 23, where scaling parameters are selected such that adding additional watermark value affects the perceived quality of the data.
25. A method of inserting a watermark into data as set forth in claim 22, where the watermark values are chosen according to a random distribution.

26. A method of inserting a watermark into data as set forth in claim 22, further comprising the step of preprocessing distorted or tampered watermarked data before said comparing data.
27. A method of inserting a watermark into data as set forth in claim 26, where said distorted or tampered watermarked data is clipped data and said preprocessing comprises replacing missing portions of the data with corresponding portions from original unwatermarked data.
28. A system for inserting a watermark into data comprising:
- providing image data;
  - providing watermark image data;
  - first transform lens for transforming image data passing therethrough into transformed image data;
  - second transform lens for transforming watermark image data passing therethrough into transformed watermark image data;
  - optical combiner for combining the transformed image data and the transformed watermark image data to form transformed watermarked data; and
  - inverse transform lens for forming watermarked data by inverse transformation of transformed watermarked data.
29. A system for inserting a watermark into data as set forth in claim 28, where said first transform lens and said second transform lens are Fourier transform lenses and said inverse transform lens is an inverse Fourier transform lens.

30. A method of inserting a watermark into data comprising the steps of:

obtaining a decomposition of data to be watermarked;

modifying the data to be watermarked by subjecting the data to distortion and/or tampering;

obtaining a decomposition of the modified data;

comparing the components of the decomposition of data to be watermarked with the components of the decomposition of the modified data; and

inserting a watermark into the data to be watermarked based upon said comparing.

### 3. Detailed Description of Invention

#### Field of the Invention

The present invention concerns a method of digital watermarking for use in audio, image, video and multimedia data for the purpose of authenticating copyright ownership, identifying copyright infringers or transmitting a hidden message. Specifically, a watermark is inserted into the perceptually most significant components of a decomposition of the data in a manner so as to be virtually imperceptible. More specifically, a narrow band signal representing the watermark is placed in a wideband channel that is the data.

#### Background of the Invention

The proliferation of digitized media such as audio, image and video is creating a need for a security system which facilitates the identification of the source of the material. The need manifests itself in terms of copyright enforcement and identification of the source of the material.

Using conventional cryptographic systems permits only valid keyholder access to encrypted data, but once the data is encrypted, it is not possible to maintain records of its subsequent representation or transmission. Conventional cryptography therefore provides minimal protection against data piracy of the type a publisher or owner of data or material is confronted with by unauthorized reproduction or distribution of such data or material.

A digital watermark is intended to complement cryptographic processes. The watermark is a visible or preferably an invisible identification code that is permanently embedded in the data. That is, the watermark remains with the data after any decryption process. As used herein the terms data and material will be understood to refer to audio (speech and music), images (photographs and graphics), video (movies or sequences of images) and

multimedia data (combinations of the above categories of materials) or processed or compressed versions thereof. These terms are not intended to refer to ASCII representations of text, but do refer to text represented as an image. A simple example of a watermark is a visible "seal" placed over an image to identify the copyright owner. However, the watermark might also contain additional information, including the identity of the purchaser of the particular copy of the image. An effective watermark should possess the following properties:

1. The watermark should be perceptually invisible or its presence should not interfere with the material being protected.
2. The watermark must be difficult (preferably virtually impossible) to remove from the material without rendering the material useless for its intended purpose. However, if only partial knowledge is known, e.g. the exact location of the watermark within an image is unknown, then attempts to remove or destroy the watermark, for instance by adding noise, should result in severe degradation in data fidelity, rendering the data useless, before the watermark is removed or lost.
3. The watermark should be robust against collusion by multiple individuals who each possess a watermarked copy of the data. That is, the watermark should be robust to the combining of copies of the same data set to destroy the watermarks. Also, it must not be possible for colluders to combine each of their images to generate a different valid watermark.
4. The watermark should still be retrievable if common signal processing operations are applied to the data. These operations include, but are not limited to digital-to-analog and analog-to-digital conversion, resampling, requantization (including dithering and recompression) and common signal enhancements to image contrast and color, or audio bass and treble for example. The watermarks in image and video data should be immune from geometric image operations such as rotation, translation, cropping and scaling.

5. The same digital watermark method or algorithm should be applicable to each of the different media under consideration. This is particularly useful in watermarking of multimedia material. Moreover, this feature is conducive to the implementation of video and image/video watermarking using common hardware.
6. Retrieval of the watermark should unambiguously identify the owner. Moreover, the accuracy of the owner identification should degrade gracefully during attack.

Several previous digital watermarking methods have been proposed. L. F. Turner in patent number WO89/08915 entitled "Digital Data Security System" proposed a method for inserting an identification string into a digital audio signal by substituting the "insignificant" bits of randomly selected audio samples with the bits of an identification code. Bits are deemed "insignificant" if their alteration is inaudible. Such a system is also appropriate for two dimensional data such as images, as discussed in an article by R.G. Van Schyndel et al entitled "A digital watermark" in Intl. Conf. on Image Processing, vol 2, Pages 86-90, 1994. The Turner method may easily be circumvented. For example, if it is known that the algorithm only affects the least significant two bits of a word, then it is possible to randomly flip all such bits, thereby destroying any existing identification code.

An article entitled "Assuring Ownership Rights for Digital Images" by G. Caronni, in Proc. Reliable IT Systems, VIS '95, 1995 suggests adding tags - small geometric patterns-to-digitized images at brightness levels that are imperceptible. While the idea of hiding a spatial watermark in an image is fundamentally sound, this scheme is susceptible to attack by filtering and redigitization. The fainter such watermarks are, the more susceptible they are to such attacks and geometric shapes provide only a limited alphabet with which to encode information. Moreover, the scheme is not applicable to audio data and may not be robust to common geometric distortions, especially cropping.

J. Brassil et al in an article entitled "Electronic Marking and Identification Techniques to Discourage Document Copying" in Proc. of Infocom 94, pp 1278-1287, 1994 propose three methods appropriate for document images in which text is common. Digital watermarks are coded by: (1)vertically shifting text lines, (2) horizontally shifting words, or (3) altering text features such as the vertical endlines of individual characters. Unfortunately, all three proposals are easily defeated, as discussed by the authors. Moreover, these techniques are restricted exclusively to images containing text.

An article by K. Tanaka et al entitled "Embedding Secret Information into a Dithered Multi-level Image" in IEEE Military Comm. Conf., pp216-220, 1990 and K. Mitsui et al in an article entitled "Video-Steganography" in IMA Intellectual Property Proc., v1, pp187-206, 1994, describe several watermarking schemes that rely on embedding watermarks that resemble quantization noise. Their ideas hinge on the notion that quantization noise is typically imperceptible to viewers. Their first scheme injects a watermark into an image by using a predetermined data stream to guide level selection in a predictive quantizer. The data stream is chosen so that the resulting watermark looks like quantization noise. A variation of this scheme is also presented, where a watermark in the form of a dithering matrix is used to dither an image in a certain way. There are several drawbacks to these schemes. The most important is that they are susceptible to signal processing, especially requantization, and geometric attacks such as cropping. Furthermore, they degrade an image in the same way that predictive coding and dithering can.

In Tanaka et al, the authors also propose a scheme for watermarking facsimile data. This scheme shortens or lengthens certain runs of data in the run length code used to generate the coded fax image. This proposal is susceptible to digital-to-analog and analog-to-digital conversions. In particular, randomizing the least significant bit (LSB) of each pixel's intensity will completely alter the resulting run length encoding. Tanaka et al also propose a watermarking method for "color-scaled picture and video sequences". This method applies the same signal transform as JPEG (DCT of 8 x 8 sub-blocks of an image) and embeds a watermark in the coefficient quantization module. While being compatible

with existing transform coders, this scheme is quite susceptible to requantization and filtering and is equivalent to coding the watermark in the least significant bits of the transform coefficients.

In a recent paper, by Macq and Quisquater entitled "Cryptology for Digital TV Broadcasting" in Proc. of the IEEE, 83(6), pp944-957, 1995 there is briefly discussed the issue of watermarking digital images as part of a general survey on cryptography and digital television. The authors provide a description of a procedure to insert a watermark into the least significant bits of pixels located in the vicinity of image contours. Since it relies on modifications of the least significant bits, the watermark is easily destroyed. Further, the method is only applicable to images in that it seeks to insert the watermark into image regions that lie on the edge of contours.

W. Bender et al in article entitled "Techniques for Data Hiding" in Proc. of SPIE, v2420, page 40, July 1995, describe two watermarking schemes. The first is a statistical method called "Patchwork". Patchwork randomly chooses  $n$  pairs of image points  $(a_i, b_i)$  and increases the brightness at  $a_i$  by one unit while correspondingly decreasing the brightness of  $b_i$ . The expected value of the sum of the differences of the  $n$  pairs of points is claimed to be  $2n$ , provided certain statistical properties of the image are true. In particular, it is assumed that all brightness levels are equally likely, that is, intensities are uniformly distributed. However, in practice, this is very uncommon. Moreover, the scheme may not be robust to randomly jittering the intensity levels by a single unit, and be extremely sensitive to geometric affine transformations.

The second method is called "texture block coding", where a region of random texture pattern found in the image is copied to an area of the image with similar texture. Autocorrelation is then used to recover each texture region. The most significant problem with this technique is that it is only appropriate for images that possess large areas of random texture. The technique could not be used on images of text, for example. Nor is there a direct analog for audio.

In addition to direct work on watermarking images, there are several works of interest in related areas. E.H. Adelson in U.S. Patent No. 4, 939,515 entitled "Digital Signal Encoding and Decoding Apparatus" describes a technique for embedding digital information in an analog signal for the purpose of inserting digital data into an analog TV signal. The analog signal is quantized into one of two disjoint ranges ( $\{0,2,4,\dots\}$ ,  $\{1,3,5\}$ , for example) which are selected based on the binary digit to be transmitted. Thus Adelson's method is equivalent to watermark schemes that encode information into the least significant bits of the data or its transform coefficients. Adelson recognizes that the method is susceptible to noise and therefore proposes an alternative scheme wherein a  $2 \times 1$  Hadamard transform of the digitized analog signal is taken. The differential coefficient of the Hadamard transform is offset by 0 or 1 unit prior to computing the inverse transform. This corresponds to encoding the watermark into the least significant bit of the differential coefficient of the Hadamard transform. It is not clear that this approach would demonstrate enhanced resilience to noise. Furthermore, like all such least significant bit schemes, an attacker can eliminate the watermark by randomization.

U.S. Patent No. 5,010,405 describes a method of interleaving a standard NTSC signal within an enhanced definition television (EDTV) signal. This is accomplished by analyzing the frequency spectrum of the EDTV signal (larger than that of the NTSC signal) and decomposing it into three sub-bands (L,M,H for low, medium and high frequency respectively). In contrast, the NTSC signal is decomposed into two subbands, L and M. The coefficients,  $M_k$ , within the M band are quantized into M levels and the high frequency coefficients,  $H_k$ , of the EDTV signal are scaled such that the addition of the  $H_k$  signal plus any noise present in the system is less than the minimum separation between quantization levels. Once more, the method relies on modifying least significant bits. Presumably, the mid-range rather than low frequencies were chosen because they are less perceptually significant. In contrast, the method proposed in the present invention modifies the most perceptually significant components of the signal.

Finally, it should be noted that many, if not all, of the prior art protocols are not collusion resistant.

Recently, Digimarc Corporation of Portland, Oregon, has described work referred to as signature technology for use in identifying digital intellectual property. Their method adds or subtracts small random quantities from each pixels. Addition or subtraction is based on comparing a binary mask of N bits with the least significant bit (LSB) of each pixel. If the LSB is equal to the corresponding mask bit, then the random quantity is added, otherwise it is subtracted. The watermark is extracted by first computing the difference between the original and watermarked images and then by examining the sign of the difference, pixel by pixel, to determine if it corresponds to the original sequence of additions/subtractions. The Digimarc technique is not based on direct modifications of the image spectrum and does not make use of perceptual relevance. While the technique appears to be robust, it may be susceptible to constant brightness offsets and to attacks based on exploiting the high degree of local correlation present in an image. For example, randomly switching the position of similar pixels within a local neighborhood may significantly degrade the watermark without damaging the image.

In a paper by Koch, Rindfrey and Zhao entitled "Copyright Protection for Multimedia Data", two general methods for watermarking images are described. The first method partitions an image into  $8 \times 8$  blocks of pixels and computes the Discrete Cosine Transform (DCT) of each of these blocks. A pseudorandom subset of the blocks is chosen and in each such block a triple of frequencies selected from one of 18 predetermined triples is modified so that their relative strengths encode a 1 or 0 value. The 18 possible triples are composed by selection of three out of eight predetermined frequencies within the  $8 \times 8$  DCT block. The choice of the eight frequencies to be altered within the DCT block appears to be based on the belief that middle frequencies have a moderate variance level, i.e., they have similar magnitude. This property is needed in order to allow the relative strength of the frequency triples to be altered without requiring a modification that would be perceptually noticeable. Unlike in the present invention, the set of frequencies is not

chosen based on any perceptual significance or relative energy considerations. In addition, because the variance between the eight frequency coefficients is small, one would expect that the technique may be sensitive to noise or distortions. This is supported by the experimental results reported in the Koch et al paper, supra, where it is reported that the "embedded labels are robust against JPEG compression for a quality factor as low as about 50%". In contrast, the method described in accordance with the teachings of the present invention has been demonstrated with compression quality factors as low as 5 percent.

An earlier proposal by Koch and Zhao in a paper entitled "Toward Robust and Hidden Image Copyright Labeling" proposed not triples of frequencies but pairs of frequencies and was again designed specifically for robustness to JPEG compression. Nevertheless, the report states that "a lower quality factor will increase the likelihood that the changes necessary to superimpose the embedded code on the signal will be noticeably visible".

In a second method, proposed by Koch and Zhao, designed for black and white images, no frequency transform is employed. Instead, the selected blocks are modified so that the relative frequency of white and black pixels encodes the final value. Both watermarking procedures are particularly vulnerable to multiple document attacks. To protect against this, Zhao and Koch proposed a distributed  $8 \times 8$  block of pixels created by randomly sampling 64 pixels from the image. However, the resulting DCT has no relationship to that of the true image. Consequently, one would expect such distributed blocks to be both sensitive to noise and likely to cause noticeable artifacts in the image.

In summary, prior art digital watermarking techniques are not robust and the watermark is easy to remove. In addition, many prior techniques would not survive common signal and geometric distortions

#### Summary of the Invention

The present invention overcomes the limitations of the prior art methods by providing a watermarking system that embeds an unique identifier into the perceptually significant components of a decomposition of an image, an audio signal or a video sequence.

Preferably, the decomposition is a spectral frequency decomposition. The watermark is embedded in the data's perceptually significant frequency components. This is because an effective watermark cannot be located in perceptually insignificant regions of image data or in its frequency spectrum, since many common signal or geometric processes affect these components. For example, a watermark located in the high frequency spectral components of an image is easily removed, with minor degradation to the image, by a process that performs low pass filtering. The issue then becomes one of how to insert the watermark into the most significant regions of the data frequency spectrum without the alteration being noticeable to an observer, i.e., a human or a machine feature recognition system. Any spectral component may be altered, provided the alteration is small. However, very small alterations are susceptible to any noise present or intentional distortion.

In order to overcome this problem, the frequency domain of the image data or sound data may be considered as a communication channel, and correspondingly the watermark may be considered as a signal transmitted through the channel. Attacks and intentional signal distortions are thus treated as noise from which the transmitted signal must be immune. Attacks are intentional efforts to remove, delete or otherwise overcome the beneficial aspects of the data watermarking. While the present invention is intended to embed watermarks in data, the same methodology can be applied to sending any type of message through media data.

Instead of encoding the watermark into the least significant components of the data, the present invention considers applying concepts of spread spectrum communication. In spread spectrum communications, a narrowband signal is transmitted over a much larger bandwidth such that the signal energy present in any single frequency is imperceptible. In

a similar manner, the watermark is spread over many frequency bins so that the energy in any single bin is small and imperceptible. Since the watermark verification process includes a priori knowledge of the locations and content of the watermarks, it is possible to concentrate these many weak signals into a single signal with a high signal to-noise ratio. Destruction of such a watermark would require noise of high amplitude to be added to every frequency bin.

In accordance with the teachings of the present invention, a watermark is inserted into the perceptually most significant regions of the data decomposition. The watermark itself is designed to appear to be additive random noise and is spread throughout the image. By placing the watermark into the perceptually significant components, it is much more difficult for an attacker to add more noise to the components without adversely affecting the image or other data. It is the fact that the watermark looks like noise and is spread throughout the image or data which makes the present scheme appear to be similar to spread spectrum methods used in communications system.

Spreading the watermark throughout the spectrum of an image ensures a large measure of security against unintentional or intentional attack. First, the location of the watermark is not obvious. Second, frequency regions are selected in a fashion that ensures severe degradation of the original data following any attack on the watermark.

A watermark that is well placed in the frequency domain of an image or a sound track will be practically impossible to see or hear. This will always be the case if the energy in the watermark is sufficiently small in any single frequency coefficient. Moreover, it is possible to increase the energy present in particular frequencies by exploiting knowledge of masking phenomena in the human auditory and visual systems. Perceptual masking refers to any situation where information in certain regions of an image or a sound is occluded by perceptually more prominent information in another part of the image or sound. In digital waveform coding, this frequency domain (and in some cases, time/pixel domain) masking is exploited extensively to achieve low bit rate encoding of data. It is clear that both

auditory and visual systems attach more resolution to the high energy, low frequency, spectral regions of an auditory or visual scene. Further, spectrum analysis of images and sounds reveals that most of the information in such data is often located in the low frequency regions.

In addition, particularly for processed or compressed data, perceptually significant need not refer to human perceptual significance, but may refer instead to machine perceptual significance, for instance, machine feature recognition.

To meet these requirements, a watermark is proposed whose structure comprises a large quantity, for instance 1000, of randomly generated numbers with a normal distribution having zero mean and unity variance. A binary watermark is not chosen because it is much less robust to attacks based on collusion of several independently watermarked copies of an image. However, generally, the watermark might have arbitrary structure, both deterministic and/or random, and including uniform distributions. The length of the proposed watermark is variable and can be adjusted to suit the characteristics of the data. For example, longer watermarks might be used for images that are especially sensitive to large modifications of its spectral coefficients, thus requiring weaker scaling factors for individual components.

The watermark is then placed in components of the image spectrum. These components may be chosen based on an analysis of those components which are most vulnerable to attack and/or which are most perceptually significant. This ensures that the watermark remains with the image even after common signal and geometric distortions. Modification of these spectral components results in severe image degradation long before the watermark itself is destroyed. Of course, to insert the watermark, it is necessary to alter these very same coefficients. However, each modification can be extremely small and, in a manner similar to spread spectrum communication, a strong narrowband watermark may be distributed over a much broader image (channel) spectrum. Conceptually, detection of the watermark then proceeds by adding all of these very small signals, whose locations are

only known to the copyright owner, and concentrating the watermark into a signal with high signal-to-noise ratio. Because the location of the watermark is only known to the copyright holder, an attacker would have to add very much more noise energy to each spectral coefficient in order to be confident of removing the watermark. However, this process would destroy the image.

Preferably, a predetermined number of the largest coefficients of the DCT (discrete cosine transform) (excluding the DC term) are used. However, the choice of the DCT is not critical to the algorithm and other spectral transforms, including wavelet type decompositions are also possible. In fact, use of the FFT rather than DCT is preferable from a computational perspective.

#### Detailed Description

In order to better understand the advantages of the invention, the preferred embodiment of a frequency spectrum based watermarking system will be described. It is instructive to examine the processing stages that image (or sound) data may undergo in the copying process and to consider the effect that such processing stages can have on the data. Referring to Figure 1, a watermarked image or sound data 10 is transmitted 12 to undergo typical distortion or intentional tampering 14. Such distortions or tampering includes lossy compression 16, geometric distortion 18, signal processing 20 and D/A and A/D conversion 22. After undergoing distortion or tampering, corrupted watermarked image or sound data 24 is transmitted 26. The process of "transmission" refers to the application of any source or channel code and/or of encryption techniques to the data. While most transmission steps are information lossless, many compression schemes (e.g., JPEG, MPEG, etc.) may potentially degrade the quality of the data through irretrievable loss of data. In general, a watermarking method should be resilient to any distortions introduced by transmission or compression algorithms.

Lossy compression 16 is an operation that usually eliminates perceptually irrelevant components of image or sound data. In order to preserve a watermark when undergoing lossy compression, the watermark is located in a perceptually significant region of the data. Most processing of this type occurs in the frequency domain. Data loss usually occurs in the high frequency components. Thus, the watermark must be placed in the

significant frequency component of the image (or sound) data spectrum to minimize the adverse affects of lossy compression.

After receipt, an image may encounter many common transformations that are broadly categorized as geometric distortions or signal distortions. Geometric distortions 18 are specific to image and video data, and include such operations as rotation, translation, scaling and cropping. By manually determining a minimum of four or nine corresponding points between the original and the distorted watermark, it is possible to remove any two or three dimensional affine transformation. However, an affine scaling (shrinking) of the image results in a loss of data in the high frequency spectral regions of the image. Cropping, or the cutting out and removal of portions of an image, also results in irretrievable loss of data. Cropping may be a serious threat to any spatially based watermark but is less likely to affect a frequency-based scheme.

Common signal distortions include digital-to-analog and analog-to-digital conversion 22, resampling, requantization, including dithering and recompression, and common signal enhancements to image contrast and/or color, and audio frequency equalization. Many of these distortions are non-linear, and it is difficult to analyze their effect in either a spatial or frequency based method. However, the fact that the original image is known allows many signal transformations to be undone, at least approximately. For example, histogram equalization, a common non-linear contrast enhancement method, may be substantially removed by histogram specification or dynamic histogram warping techniques.

Finally, the copied image may not remain in digital form. Instead, it is likely to be printed or an analog recording made (analog audio or video tape). These reproductions introduce additional degradation into the image data that a watermarking scheme must be robust to.

Tampering ( or attack) refers to any intentional attempt to remove the watermark, or corrupt it beyond recognition. The watermark must not only be resistant to the inadvertent application of distortions. It must also be immune to intentional manipulation

by malicious parties. These manipulations can include combinations of distortions, and can also include collusion and forgery attacks.

Figure 2 shows a preferred system for inserting a watermark into an image in the frequency domain. Image data  $X(i,j)$  assumed to be in digital form, or alternatively data in other formats such as photographs, paintings or the like, that have been previously digitized by well-known methods, is subject to a frequency transformation 30, such as the Fourier transform. A watermark signal  $W(k)$  is inserted into the frequency spectrum components of the transformed image data 32 applying the techniques described below. The frequency spectrum image data including the watermark signal is subjected to an inverse frequency transform 34, resulting in watermarked image data  $\hat{X}(i,j)$ , which may remain in digital form or be printed as an analog representation by well-known methods.

After applying a frequency transformation to the image data 30, a perceptual mask is computed that highlights prominent regions in the frequency spectrum capable of supporting the watermark without overly affecting perceptual fidelity. This may be performed by using knowledge of the perceptual significance of each frequency in the spectrum, as discussed earlier, or simply by ranking the frequencies based on their energy. The latter method was used in experiments described below.

In general, it is desired to place the watermark in regions of the spectrum that are least affected by common signal distortions and are most significant to image quality as perceived by a viewer, such that significant modification would destroy the image fidelity. In practice, these regions could be experimentally identified by applying common signal distortions to images and examining which frequencies are most affected, and by psychophysical studies to identify how much each component may be modified before significant changes in the image are perceivable.

The watermark signal is then inserted into these prominent regions in a way that makes any tampering create visible (or audible) defects in the data. The requirements of the

watermark mentioned above and the distortions common to copying provide constraints on the design of an electronic watermark.

In order to better understand the watermarking method, reference is made to Figures 3(a) and 3(b) where from each document D a sequence of values  $X=x_1, \dots, x_n$  is extracted 40 with which a watermark  $W=w_1, \dots, w_n$  is combined 42 to create an adjusted sequence of values  $X'=x'_1, \dots, x'_n$  which is then inserted back 44 into the document in place of values X in order to obtain a watermark document D'. An attack of the document D', or other distortion, will produce a document D\*. Having the original document D and the document D\*, a possibly corrupted watermark W\* is extracted 46 and compared to watermark W 48 for statistical analysis 50. The values W\* are extracted by first extracting a set of values  $X^*=x_1^*, \dots, x_n^*$  from D\* (using information about D) and then generating W\* from the values X\* and the values X.

When combining the values X with the watermark values W in step 42, scaling parameter  $\alpha$  is specified. The scaling parameter  $\alpha$  determines the extent to which values W alter values X. Three preferred formulas for computing X' are:

$$x'_i = x_i + \alpha w_i \quad (1)$$

$$x'_i = x_i (1 + \alpha w_i) \quad (2)$$

$$x'_i = x_i (e^{\alpha w_i}) \quad (3)$$

Equation 1 is invertible. Equations 2 and 3 are invertible when  $x_i \neq 0$ . Therefore, given X\* it is possible to compute the inverse function necessary to derive W\* from X and X\*.

Equation 1 is not the preferred formula when the values  $x_i$  vary over a wide range. For example, if  $x_i=10^6$  then adding 100 may be insufficient to establish a watermark, but if  $x_i=10$ , then adding 100 will unacceptably distort the value. Insertion methods using equations 2 and 3 are more robust when encountering such a wide range of values  $x_i$ . It

will also be observed that equation 2 and 3 yield similar results when  $\alpha w_i$  is small. Moreover, when  $x_i$  is positive, equation 3 is equivalent to  $\ln(x_i) = \ln(x_i) + \alpha x_i$  and may be considered as an application of equation 1 when natural logarithms of the original values are used. For example, if  $|w_i| \leq 1$  and  $\alpha = 0.01$ , then using Equation (2) guarantees that the spectral coefficient will change by no more than 1%.

For certain applications, a single scaling parameter  $\alpha$  may not be best for combining all values of  $x_i$ . Therefore, multiple scaling parameters  $\alpha_1, \dots, \alpha_n$  can be used with revised equations 1 to 3 such as  $x_i = x_i (1 + \alpha_i w_i)$ . The values of  $\alpha_i$  serve as a relative measure of how much  $x_i$  must be altered to change the perceptual quality of the document. A large value for  $\alpha_i$  means that it is possible to alter  $x_i$  by a large amount without perceptually degrading the document.

A method for selecting the multiple scaling values is based upon certain general assumptions. For example, equation 2 is a special case of the generalized equation 1,  $(x_i' = x_i + \alpha_i x_i)$ , for  $\alpha_i = \alpha w_i$ . That is, equation 2 makes the reasonable assumption that a large value of  $x_i$  is less sensitive to additive alteration than a small value of  $x_i$ .

Generally, the sensitivity of the image to different values of  $\alpha_i$  is unknown. A method of empirically estimating the sensitivities is to determine the distortion caused by a number of attacks on the original image. For example, it is possible to compute a degraded image  $D^*$  from  $D$ , extract the corresponding values  $x_1^*, \dots, x_n^*$  and select  $\alpha_i$  to be proportional to the deviation  $|x_i^* - x_i|$ . For greater robustness, it is possible to try other forms of distortion and make  $\alpha_i$  proportional to the average value of  $|x_i^* - x_i|$ . Instead of using the average deviation, it is possible to use the median or maximum deviation.

Alternatively, it is possible to combine the empirical approach with general global assumptions regarding the sensitivity of the values. For example, it might be required that

$\alpha_i \geq \alpha_j$  whenever  $x_i \geq x_j$ . This can be combined with the empirical approach by setting  $\alpha_i$  according to

$$\alpha_i \sim \max_{\{j | v_j \leq v_i\}} |v_j^* - v_j|$$

A more sophisticated approach is to weaken the monotonicity constraint to be robust against occasional outliers.

The length of the watermark,  $n$ , determines the degree to which the watermark is spread among the relevant components of the image data. As the size of the watermark increases, so does the number of altered spectral components, and the extent to which each component need be altered decreases for the same resilience to noise. Consider watermarks of the form  $x_i' = x_i + \alpha w_i$  and a white noise attack by  $x_i' = x_i' + r_i$  where  $r_i$  are chosen according to independent normal distributions with standard deviation  $\sigma$ . It is possible to recover the watermark when  $\alpha$  is proportional to  $\sigma/\sqrt{n}$ . That is, quadrupling the number of components can halve the magnitude of the watermark placed into each component. The sum of the squares of the deviations remains essentially unchanged.

In general, a watermark comprises an arbitrary sequence of real numbers  $W = w_1, \dots, w_n$ . In practice, each value  $w_i$  may be chosen independently from a normal distribution  $N(0,1)$ , where  $N(\mu, \sigma^2)$  with mean  $\mu$  and variance  $\sigma^2$  or of a uniform distribution from  $(-1,1)$  or  $(0,1)$ .

It is highly unlikely that the extracted mark  $W^*$  will be identical to the original watermark  $W$ . Even the act of requantizing the watermarked document for transmission will cause  $W^*$  to deviate from  $W$ . A preferred measure of the similarity of  $W$  and  $W^*$  is

$$\text{sim}(W, W^*) = \frac{W^* \cdot W}{\sqrt{W^* \cdot W^*}} \quad (4)$$

Large values of  $\text{sim}(W, W^*)$  are significant in view of the following analysis. Assume that the authors of document  $D^*$  had no access to  $W$  (either through the seller or through a

watermarked document). Then for whatever value of  $W^*$  is obtained, the conditional distribution on  $w_i$  will be independently distributed according to  $N(0,1)$ . In this case,

$$N(0, \sum_{i=1}^n x_i^2) = N(0, W^* \cdot W^*).$$

Thus,  $\text{sim}(W, W^*)$  is distributed according to  $N(0,1)$ . Then, one may apply the standard significance tests for the normal distribution. For example, if  $D^*$  is chosen independently from  $W$ , then it is very unlikely that  $\text{sim}(W, W^*) > 5$ . Note that somewhat higher values of  $\text{sim}(W, W^*)$  may be needed when a large number of watermarks are on file. The above analysis required only the independence of  $W$  from  $W^*$ , and did not rely on any specific properties of  $W^*$  itself. This fact provides further flexibility when preprocessing  $W^*$ .

The extracted watermark  $W^*$  may be extracted in several ways to potentially enhance the ability to extract a watermark. For example, experiments on images encountered instances where the average value of  $W^*$ , denoted  $E_i(W^*)$ , differed substantially from 0, due to the effects of a dithering procedure. While this artifact could be easily eliminated as part of the extraction process, it provides a motivation for postprocessing extracted watermarks. As a result, it was discovered that the simple transformation  $w_i^* \leftarrow w_i^* - E_i(W^*)$  yielded superior values of  $\text{sim}(W, W^*)$ . The improved performance resulted from the decreased value of  $W^* \cdot W^*$ ; the value of  $W^* \cdot W$  was only slightly affected.

In experiments it was frequently observed that  $w_i^*$  could be greatly distorted for some values of  $i$ . One postprocessing option is to simply ignore such values, setting them to 0. That is,

$$w_i^* \leftarrow \begin{cases} w_i^* & \text{if } |w_i^*| > \text{tolerance} \\ 0 & \text{otherwise} \end{cases}$$

The goal of such a transformation is to lower  $W^* \cdot W^*$ . A less abrupt version of this approach is to normalize the  $W^*$  values to be either -1, 0 or 1, by

$$w_i^* \leftarrow \text{sign}(w_i^* - E_i(W^*)).$$

This transformation can have a dramatic effect on the statistical significance of the result. Other robust statistical techniques could also be used to suppress outlier effects.

In principle, any frequency domain transform can be used. In the scheme described below, a Fourier domain method is used, but the use of wavelet based schemes are also useable as a variation. In terms of selecting frequency regions of the transform, it is possible to use models for the perceptual system under consideration.

Frequency analysis may be performed by a wavelet or sub-band transform where the signal is divided into sub-bands by means of a wavelet or multi-resolution transform. The sub-bands need not be uniformly spaced. Each sub-band may be thought of as representing a frequency region in the domain corresponding to a sub-region of the frequency range of the signal. The watermark is then inserted into the sub-regions.

For audio data, a sliding "window" moves along the signal data and the frequency transform (DCT, FFT, etc.) is taken of the sample in the window. This process enables the capture of meaningful information of a signal that is time varying in nature.

Each coefficient in the frequency domain is assumed to have a perceptual capacity. That is, it can support the insertion of additional information without any (or with minimal) impact to the perceptual fidelity of the data.

In order to place a length L watermark into an  $N \times N$  image, the  $N \times N$  FFT (or DCT) of the image is computed and the watermark is placed into the L highest magnitude coefficients of the transform matrix, excluding the DC component. More generally, L randomly chosen coefficients could be chosen from the M,  $M \geq L$  most perceptually significant coefficients of the transform. For most images, these coefficients will be the ones corresponding to the low frequencies. The purpose of placing the watermark in

these locations is because significant tampering with these frequencies will destroy the image fidelity or perceived quality well before the watermark is destroyed.

The FFT provides perceptually similar results to the DCT. This is different than the case of transform coding, where the DCT is preferred to the FFT due to its spectral properties. The DCT tends to have less high frequency information than that the FFT, and places most of the image information in the low frequency regions, making it preferable in situations where data need to be eliminated. In the case of watermarking, image data is preserved, and nothing is eliminated. Thus the FFT is as good as the DCT, and is preferred since it is easier to compute.

In an experiment, a visually imperceptible watermark was intentionally placed in an image. Subsequently, 100 randomly generated watermarks, only one of which corresponded to the correct watermark, were applied to the watermark detector described above. The result, as shown in Figure 4, was a very strong positive response corresponding to the correct watermark, suggesting that the method results in a very low number of false positive responses and a very low false negative response rate.

In another test, the watermarked image was scaled to half of its original size. In order to recover the watermark, the image was re-scaled to its original size, albeit with loss of detail due to subsampling of the image using low pass spatial filter operations. The response of the watermark detector was well above random chance levels, suggesting that the watermark is robust to geometric distortions. This result was achieved even though 75 percent of the original data was missing from the scaled down image.

In a further experiment, a JPEG encoded version of the image with parameters of 10 percent quality and 0 percent smoothing, resulting in visible distortions, was used. The results of the watermark detector suggest that the method is robust to common encoding distortions. Even using a version of the image with parameters of the 5 percent quality

and 0 percent smoothing, the results were well above that achievable due to random chance.

In experiments using a dithered version of the image, the response of the watermark detector suggested that the method is robust to common encoding distortion. Moreover, more reliable detection is achieved by removing any non-zero mean from the extracted watermark.

In another experiment, the image was clipped, leaving only the central quarter of the image. In order to extract the watermark from the clipped image, the missing portion of the image was replaced with portions from the original unwatermarked image. The watermark detector was able to recover the watermark with a response greater than random. When the non-zero mean was removed, and the elements of the watermark were binarized prior to the comparison with the correct watermark, the detector response was improved. This result is achieved even though 75 percent of the data was removed from the image.

In yet another experiment, the image was printed, photocopied, scanned using a 300 dpi Umax PS-2400x scanner and rescaled to a size of  $256 \times 256$  pixels. Clearly, the final image suffered from different levels of distortion introduced at each process. High frequency pattern noise was particularly noticeable. When the non-zero mean was removed and only the sign of the elements of the watermark was used, the watermark detector response improved to well above random chance levels.

In still another experiment, the image was subject to five successive watermarking operations. That is, the original image was watermarked, the watermarked image was watermarked, and so forth. The process may be considered another form of attack in which it is clear that significant image degradation occurs if the process is repeated. Figure 5 shows the response of the watermark detector to 1000 randomly generated watermarks, including the five watermarks present in the image. The five dominant spikes

in the graph, indicative of the presence of the five watermarks, show that successive watermarking does not interfere with the process.

The fact that successive watermarking is possible means that the history or pedigree of a document is determinable if successive watermarking is added with each copy.

In a variation of the multiple watermark image, five separately watermarked images were averaged together to simulate simple collusion attack. Figure 6 shows the response of the watermark detector to 1000 randomly generated watermarks, including the five watermarks present in the original images. The result is that simple collusion based on averaging is ineffective in defeating the present watermarking system.

The result of the above experiments is that the described system can extract a reliable copy of the watermark from images that have been significantly degraded through several common geometric and signal processing procedures. These procedures include zooming (low pass filtering), cropping, lossy JPEG encoding, dithering, printing, photocopying and subsequent rescanning.

While these experiments were, in fact, conducted using an image, similar results are attainable with text images, audio data and video data, although attention must be paid to the time varying nature of these data.

The above implementation of the watermarking system is an electronic system. Since the basic principle of the invention is the inclusion of a watermark into spectral frequency components of the data, watermarking can be accomplished by other means using, for example, an optical system as shown in Figure 7.

In Figure 7, data to be watermarked such as an image 40 is passed through a spatial transform lens 42, such as a Fourier transform lens, the output of which lens is the spatial transform of the image. Concurrently, a watermark image 44 is passed through a second

spatial transform lens 46, the output of which lens is the spatial transfer of the watermark image 44. The spatial transform from lens 42 and the spatial transform from lens 46 are combined at an optical combiner 48. The output of the optical combiner 48 is passed through an inverse spatial transform lens 50 from which the watermark image 52 is present. The result is a unique, virtually imperceptible, watermarked image. Similar results are achievable by transmitting video or multimedia signals through the lenses in the manner described above.

While there have been described and illustrated spread spectrum watermarking of data and variations and modifications thereof, it will be apparent to those skilled in the art that further variations and modifications are possible without deviating from the broad principles and spirit of the present invention which shall be limited solely by the scope of the claims appended hereto.

#### 4. Brief Description of Drawings

**Figure 1** is a schematic representation of typical common processing operations to which data could be subjected;

**Figure 2** is a schematic representation of a preferred system for immersing a watermark into an image;

**Figures 3a and 3b** are flow charts of the encoding and decoding of watermarks;

**Figure 4** is a graph of the responses of the watermark detector to random watermarks;

**Figure 5** is a graph of the response of the watermark detector to random watermarks for an image which is successively watermarked five times;

**Figure 6** is a graph of the response of the watermark detector to random watermarks where five images, each having a different watermark, and averaged together; and

**Figure 7** is a schematic diagram of an optical embodiment of the present invention

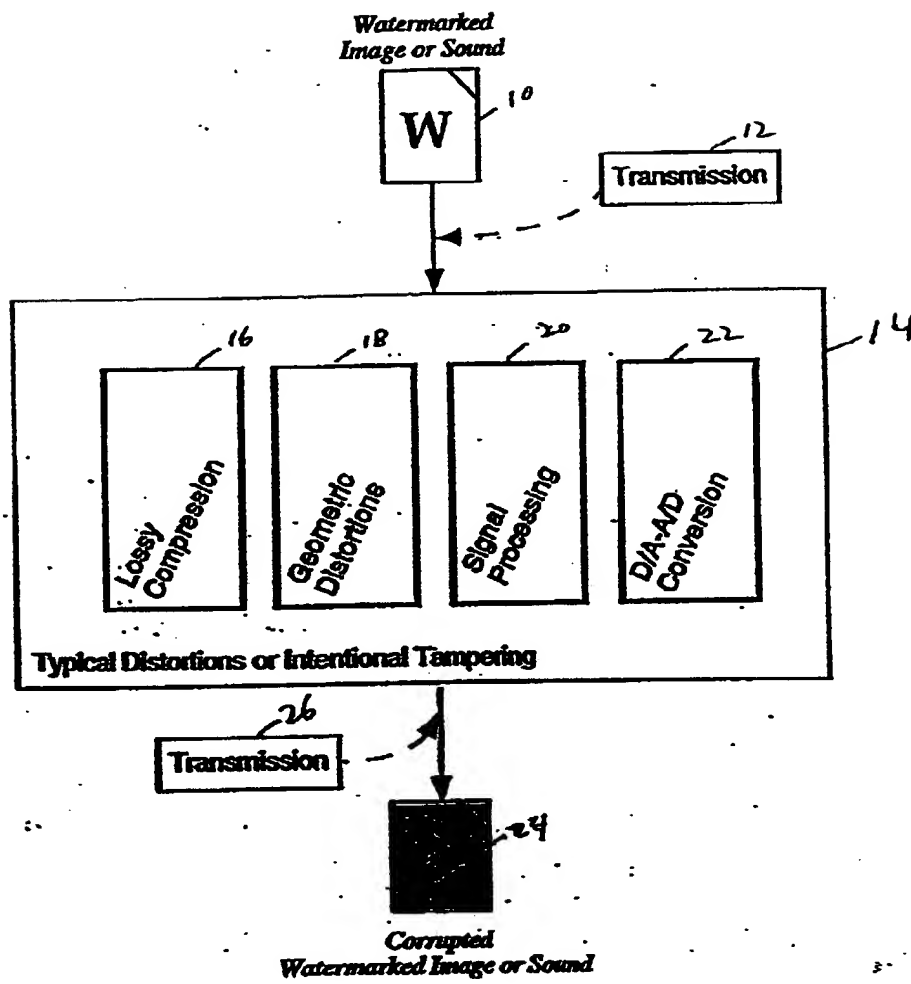


Figure 1

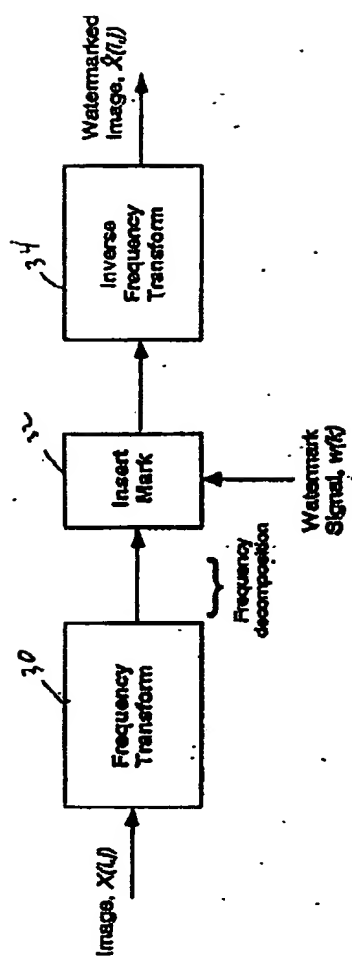


Figure 2

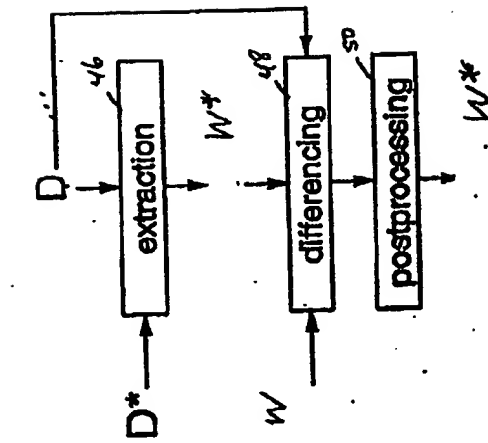


FIGURE 3(a)

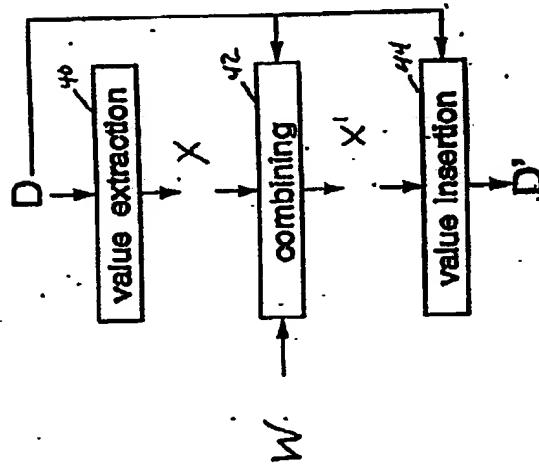


FIGURE 3(b)

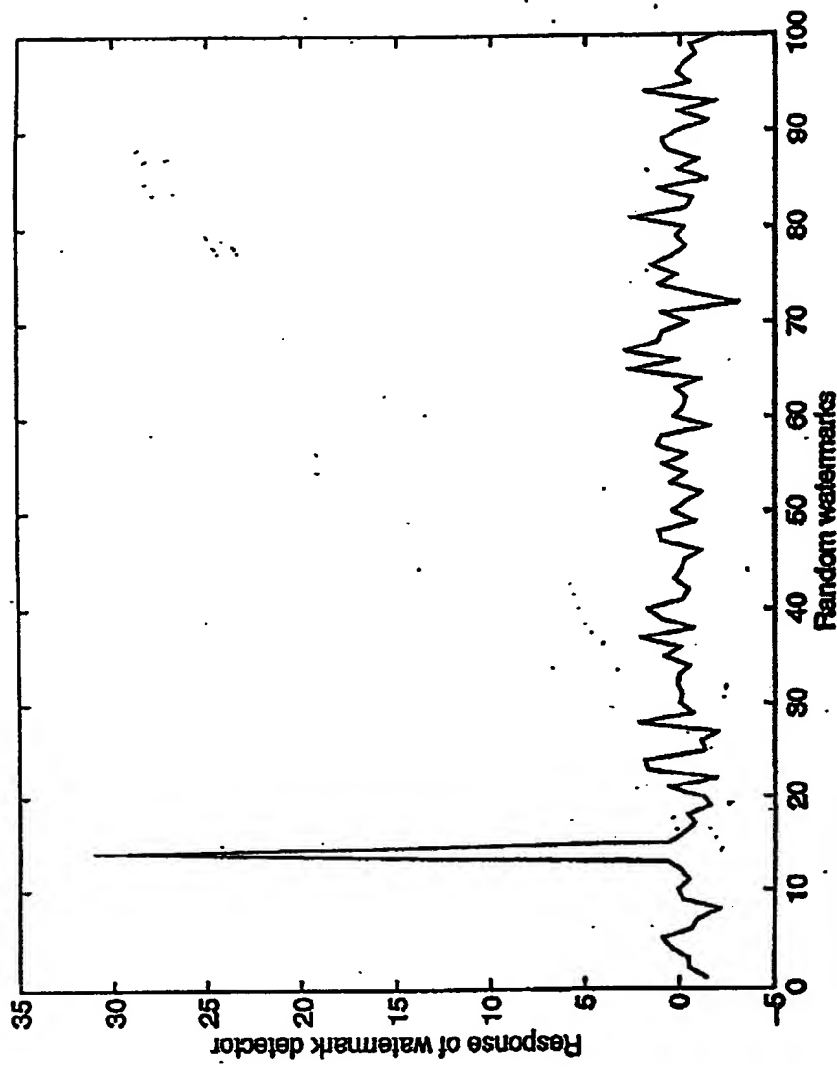


FIGURE 4

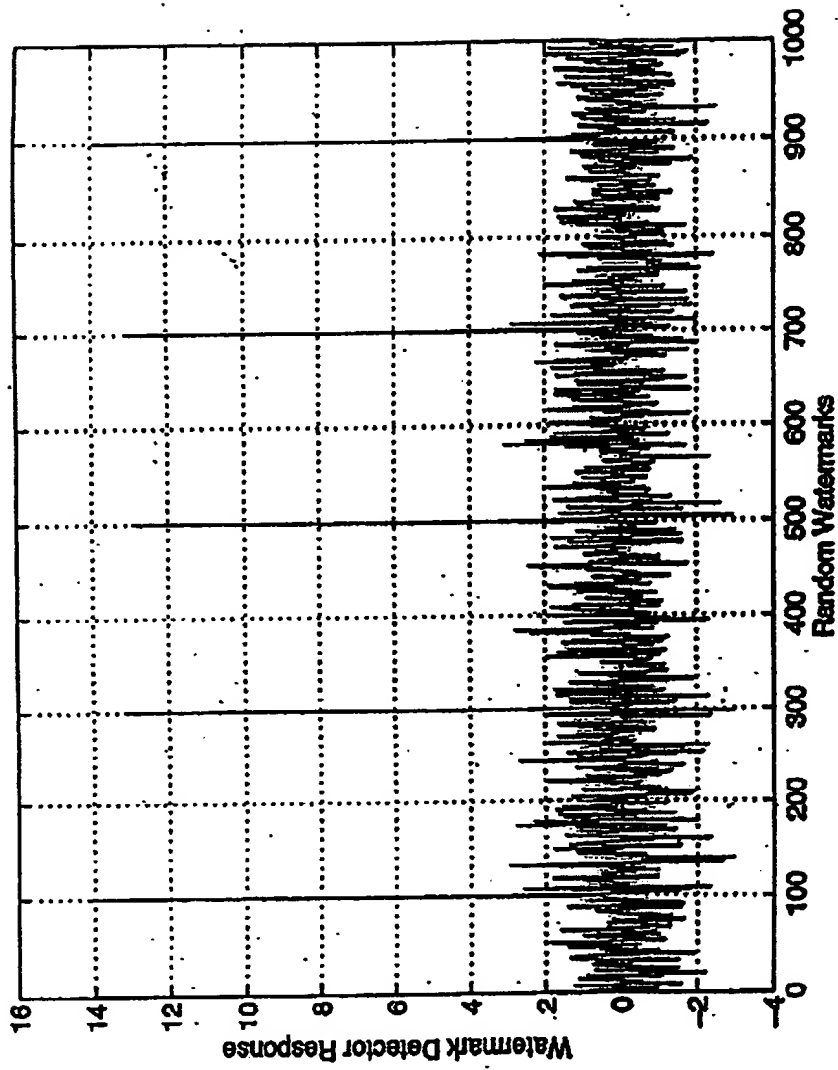


FIGURE 5

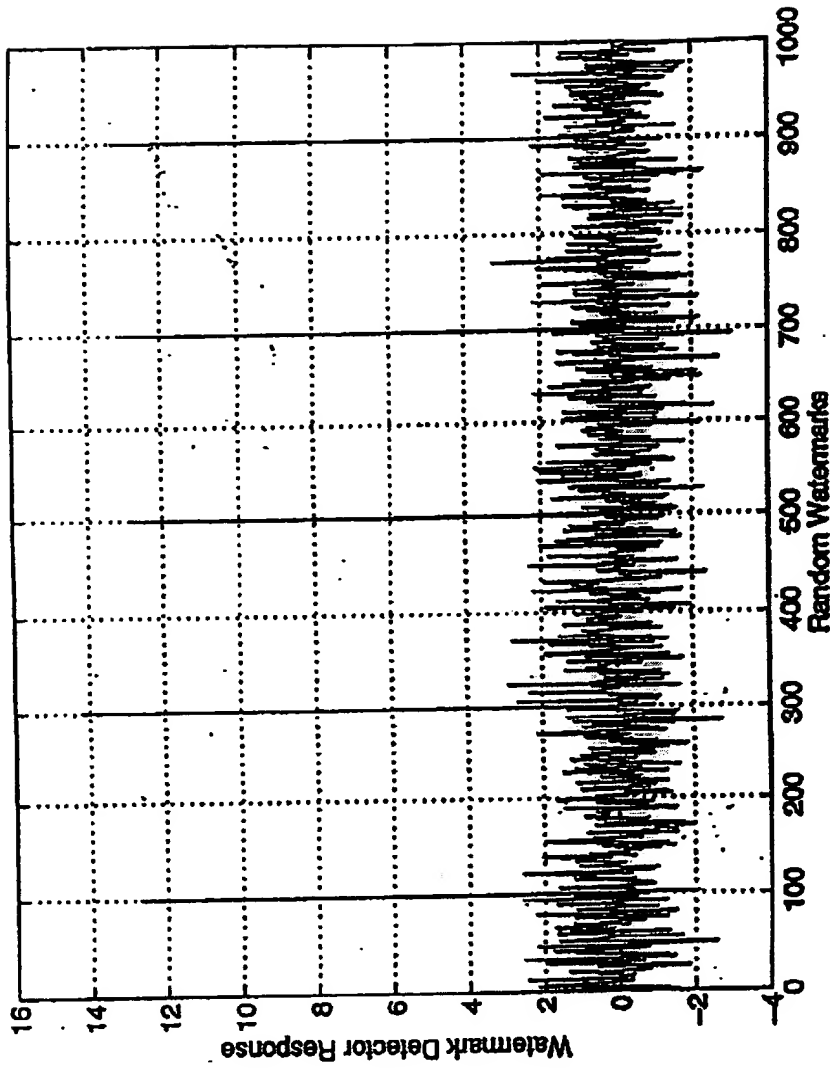


FIGURE 6

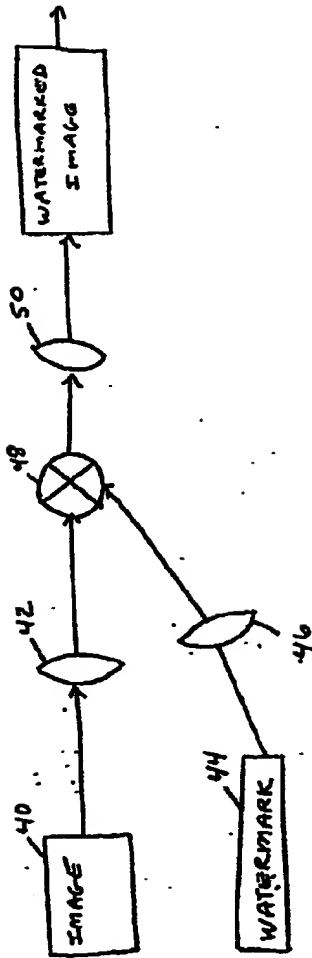


FIGURE 7

## 1. Abstract

Digital watermarking of audio, image, video or multimedia data is achieved by inserting the watermark into the perceptually significant components of a decomposition of the data in a manner so as to be visually imperceptible. In a preferred method, a frequency spectral image of the data, preferably a Fourier transform of the data, is obtained. A watermark is inserted into perceptually significant components of the frequency spectral image. The resultant watermarked spectral image is subjected to an inverse transform to produce watermarked data. The watermark is extracted from watermarked data by first comparing the watermarked data with the original data to obtain an extracted watermark. Then, the original watermark, original data and the extracted watermark are compared to generate a watermark which is analyzed for authenticity of the watermark.

## 2. Representative Drawing

Figure 2